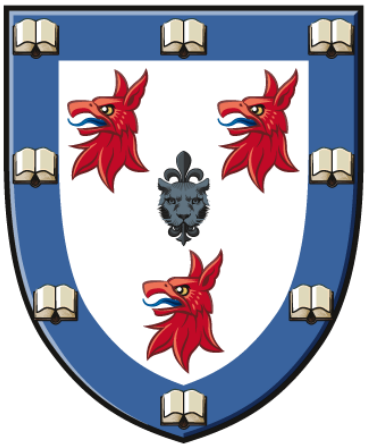
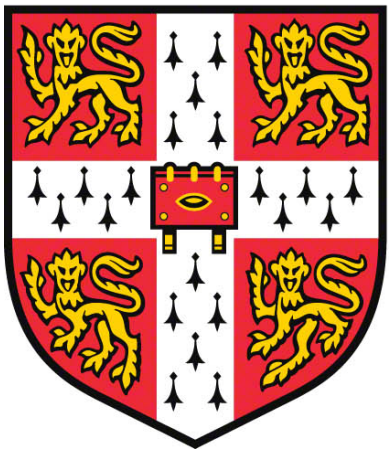
Probing Order within Intrinsically Disordered Proteins





Michael David Crabtree

Homerton College

University of Cambridge

September 2017

This dissertation is submitted for the degree of

Doctor of Philosophy

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution, except as declared in the Preface and specified in the text. It does not exceed the prescribed 60,000-word limit imposed by the Physics and Chemistry Degree Committee.

Michael David Crabtree

September 2017

Summary

Decades have passed since the realisation that a protein's amino acid sequence can contain all the information required to form a complex three-dimensional fold. Until recently, these encoded structures were thought to be crucial determinants of protein function. Much effort was directed to fully understand the mechanisms behind how and why proteins fold, with natively unfolded proteins thought to be experimental artefacts. Today, the field of natively unfolded – or so-called intrinsically disordered – proteins, is rapidly developing. Protein disorder content has been positively correlated with organismal complexity, with over thirty percent of eukaryotic proteins predicted to contain disordered regions. However, the biophysical consequences of disorder are yet to be fully determined. With the aim of addressing some of the outstanding questions, the work described in this thesis focuses on the relevance of structure within disordered proteins.

Whilst populating a variety of conformations in isolation, a subset of disordered proteins can fold upon binding to a partner macromolecule. This folded state may be present within the ensemble of conformations sampled by the unbound protein, opening the question of what comes first: folding or binding? Protein engineering techniques were employed to alter the level of residual 'bound-like' structure within the free conformational ensemble, and the consequences on coupled folding and binding reactions were investigated. Resultant changes in the rate of association are easily imaginable; yet, this work demonstrates that the majority of the observed changes in binding affinity were due to alterations in the rate of dissociation, thus altering the lifetime of the bound complex.

Promiscuous binding is a touted advantage of being disordered. If many disordered proteins, each with their own conformational ensemble, can bind and fold to the same partner, then where is the folding component encoded? Does the partner protein template the folding reaction? Or, is the folding information contained within the disordered protein sequence? Utilising phi-value analysis on the BCL-2 family of proteins, residues in the disordered sequence were probed to ascertain which form contacts at the transition state of the reaction. Comparison with phi-value analyses of alternative pairs – sharing either the ordered or disordered protein – provides insight into the encoding of these interactions. In the context of a bimolecular reaction, the amino acid sequence of the disordered protein was shown to determine the interactions within the transition state. Thus, analogous to the

discovery from decades' past, it is the sequence of the protein that folds which encodes its pathway, even when binding is a prerequisite.

Acknowledgements

First of all, I would like to thank my parents. Throughout my entire life they have encouraged, supported and nurtured me. I am forever grateful for all the time and effort that they devote to me. Without their guidance, I do not think I would be the person I am today.

At times, these last four years have been hard. Long days, experimental worries and equipment failures have, on occasion, soured my mood. I must thank my partner, Lauren for constantly supporting me, cheering me up and making me realise that it will all be OK.

A huge number of people have helped me throughout both my master's degree and PhD. I would like to thank everyone in Clarke Group and Lab 290 for everything they have done for me. In particular, I would like to thank Jane for welcoming me into her lab. Ever since my first day, she has been a great mentor, continually encouraging, guiding and pushing me to do better.

When I first arrived in Clarke Group, I somehow managed to know even less than I do now. Dazzled by the equipment and terminology, I was like a rabbit stuck in headlights. Fortunately, the group members at the time: Sarah; Katie; Alex; Dominika; Joe; Carol; Stephanie; Adrian and Andrew, took me under their wing and I learnt a great deal from them. In particular, I would like to thank Sarah. She supervised me at the start of my PhD, teaching me biophysical techniques and scientific rigour. I hope that the work presented in this thesis lives up to her fantastic training.

Over the years, many people have come and gone in Clarke Group. While all of them have positively impacted on my life, there are a few that I would like to specifically mention. I started my PhD at roughly the same time as a few other students: Basile; Liza; Carol and Sophie. Although we have all been on this journey together, there has been one person who, due to their proximity, has had to deal with me the most. Basile has been my desk partner for almost three years. His scientific knowledge and intelligence have driven me to think further and aim higher. I do not think that I would have understood as much about science or achieved as much as I have, if it was not for our - sometimes heated - discussions. I would also like to thank Carol, Liza, Tristan, Jeff and Andrew. All of them are good

friends, and have regularly helped me. In the early days, I worked closely with Liza. Her questioning of my assumptions encouraged me to delve deeper into the theory behind our experiments. In the last year, I have transitioned to working more closely with Carol. The bidirectional passing of knowledge between us has been a delight – I am particularly grateful for the time she spent teaching me how to perform chemical denaturant experiments. As part of the project with Carol, I supervised a master's student, Quenton. During his stint in the lab, Quenton achieved a great deal, and I would like to thank him for the time and effort he put into making the project a success. Another member of Clarke group has specifically made my final year a lot easier and more productive. Tristan has made countless protein preps at short notice, helping me to complete projects that otherwise may have been left unfinished.

I would also like to thank Anita, Jo, Helen, Annette and Ryan. Chats with Anita have brightened many a morning, while her kind preparation of media has saved me countless hours. Several experiments in this thesis required equipment from the Centre for Biological Chemistry. During his time in the department, Ryan helped me troubleshoot problems with the CBC equipment on innumerable occasions. Through kindly giving me their time and assisting me with unfathomably complex admin tasks, Jo and Helen have generally made my life a lot easier. Annette has been particularly helpful during this writing up period, proof reading all of my chapters.

The work described in several chapters of this thesis would not have been the same without the help of our collaborators. I would like to thank Gary Daughdrill, Wade Borcherds, Anusha Poosapati, Rohit Pappu, Tyler Harmon and Ammon Posey for their contributions. Additionally, much of the work required accurate knowledge of protein concentrations and identity. This was achieved thanks to Peter Sharratt and Asha Boodhun, who performed amino acid analysis and mass spectrometry, respectively.

Finally, I would like to thank my family and everyone else who has supported me. I am extremely grateful to all of you.

Abbreviations

General

‡	Transition State
D	Denatured State
GADIS	Genetic Algorithm for Design of Intrinsic Secondary Structure
K_d	Equilibrium Dissociation Constant
k_{obs}	Observed Association Rate Constant
k_{off}	Dissociation Rate Constant
k_{on}	Association Rate Constant
Mut	Mutant
N	Native State
RMSD	Root-Mean-Square Deviation
TNF	Tumour Necrosis Factor
WT	Wild-Type

Reagents

2x YT	2x Yeast Tryptone Broth
ddH_2O	Ultrapure Deionised Water
DTT	Dithiothreitol
GdmCl	Guanidine Hydrochloride
IPTG	isopropyl β-D-Thiogalactopyranoside
LB	Lysogeny Broth
TCEP	Tris(2-Carboxyethyl)Phosphine

Techniques

CD	Circular Dichroism
FPLC	Fast Protein Liquid Chromatography
ITC	Isothermal Titration Calorimetry
NMR	Nuclear Magnetic Resonance
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

Proteins

BCL-2	B-Cell Lymphoma 2
c-MYB	Transcriptional Activator MYB
CBP-KIX	CREB-Binding Protein-Kinase Inducible Domain Interacting Domain
MCL-1	Induced Myeloid Leukaemia Cell Differentiation Protein MCL-1
MDM2	E3 Ubiquitin-Protein Ligase MDM2
MLL	Mixed Lineage Leukaemia
PUMA	p53 Upregulated Modulator of Apoptosis

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Human Genome Project

Deoxyribonucleic acid (DNA): words that many people may recognise, perhaps without grasping the biological importance or function that the term entails. Enrolled in a media storm, DNA shot to prominence at the start of the 21st century. Scientists from around the world were both collaborating and competing to publish the sequence of nucleotides that represent the human instruction manual. Within the same week in 2001, both an academic conglomerate and a private company presented their findings (International Human Genome Sequencing Consortium, 2001; Venter *et al.*, 2001). The race was over. The draft of the human genome was published.

Such was the perceived importance of the human genome project, Presidents and Prime Ministers had, 8-months before publication of the papers, pre-emptively hailed the benefits of knowing humans' genetic code. The perceived advantages are as a result of DNA's significance in biology. DNA is a polynucleotide, consisting of 4 nucleotide bases; adenosine (a), thymine (t), guanine (g) and cytosine (c). When read in groups of three, these nucleotides contain the coding information for an amino acid residue (e.g. ccc encodes proline). The name attached to these groups of three nucleotides, a codon, references the data storing ability of DNA. However, on its own, a codon only provides information for a single amino acid. The real power of DNA comes through combining codons into genes. Genes are a series of codons that encode multiple, covalently linked amino acids. These amino acid polymers, termed proteins, are key players in biology.

Much of the hype surrounding the Human Genome Project related to the promise that it would revolutionise medicine. Diseases can contain a genetic aspect, where differences in the nucleotide sequence from one individual to another can either directly cause, or make one more susceptible to, a specific illness. For example, in the gene for the cystic fibrosis transmembrane conductance regulator (CFTR),

deletion of a single codon can cause the disease, cystic fibrosis (Kerem *et al.*, 1989; Riordan *et al.*, 1989; Rommens *et al.*, 1989). However, knowledge of the genetic causes does not automatically result in the disease being cured. The true potential of the Human Genome Project related to identifying which proteins were involved in disease, and how they differed from the healthy population. This point is illustrated by returning to the cystic fibrosis example mentioned earlier. Rather than directly causing the disease, the codon deletion in the CFTR gene results in the deletion of an amino acid (phenylalanine) at position 508. This leads to higher rates of degradation, effectively meaning that the protein is no longer present (O'Sullivan and Freedman, 2009), and the protein function is lost. Thus, while DNA provides the starting point for what has gone wrong, understanding how and why the protein function has changed is required in order to fully understand the disease biology.

1.2 Amino Acids

Given the requirement for codons, the 4 DNA nucleotide bases could encode a maximum of 64 amino acids (4^3). If this were the case, then the potential for problematic changes would be vast: mutation of a single nucleotide would result in a different amino acid, potentially altering the function of the protein. Instead, biology utilises just 20 amino acids, allowing several codons to encode the same residue. Based on their chemical characteristics, amino acids can be divided into different groups (Figure 1.1): non-polar; aromatic; polar; negatively charged; and positively charged.

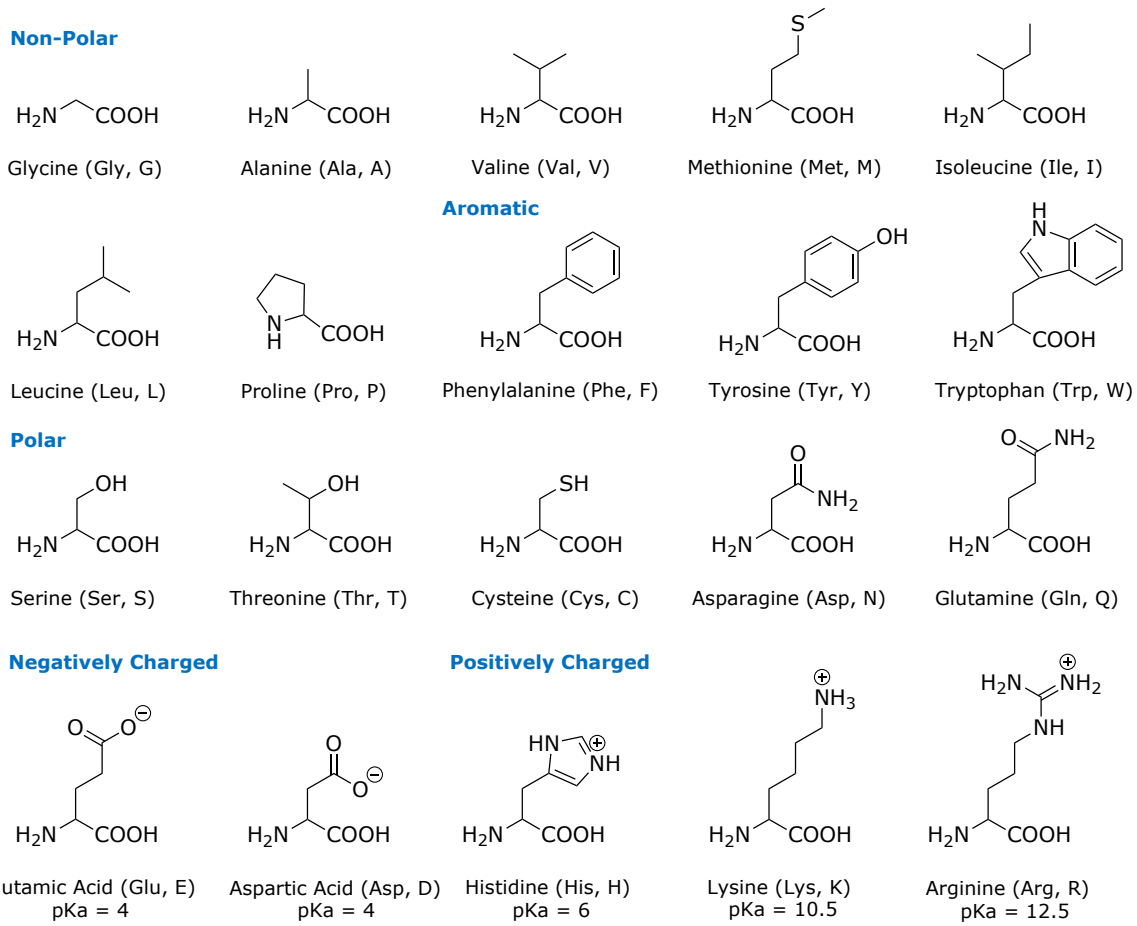


Figure 1.1 Amino Acid Structures and Properties.
The full name, three letter and single letter codes are shown for each of the 20 standard amino acids. Amino acid side chains from the alpha-carbon vary in size and property. For the charged residues, approximate pKa values are shown, and the side chain charge of the majority species at pH 7.0 is indicated. Structures were constructed using ChemDraw Professional (version 16.0).

To form a protein, the carboxyl group of one amino acid undergoes a condensation reaction with the amino group of another residue. The resultant peptide bond (ω) is planar and has a relatively fixed angle of either 180° (*trans*) or 0° (*cis*) (Figure 1.2). Steric clashes between the side chains, and high energy barriers limit the population of peptide bonds found in *cis*, with more than 99% typically observed in the *trans* conformation (Stewart *et al.*, 1990). At least 2 rotatable bonds are present in each amino acid – the Cα-C (psi torsion angle, ψ) and the N-Cα (phi torsion angle, φ) bonds. Additional rotational bonds are provided by the various amino acid side chains (Figure 1.1).

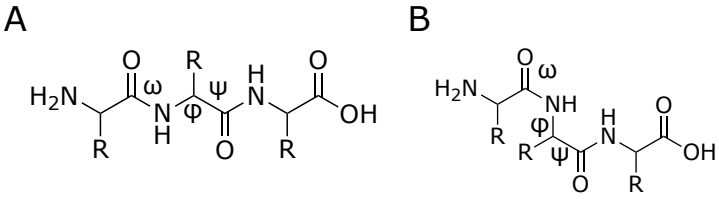


Figure 1.2 Peptide Bond.
A tripeptide with the first peptide bond (ω) shown in either *trans* (A) or *cis* (B). The phi (φ) and psi (ψ) torsional angles are also indicated. R indicates amino acid side chains. Structures were constructed using ChemDraw Professional (version 16.0).

1.3 Forces that Fold

In biology, proteins are typically found in solution. Interactions between the protein and solvent can alter the phi and psi angles, changing the conformation of the protein. Depending on the characteristics of the side chain and the solvent, these interactions can either be favourable or unfavourable. Forces between amino acids can also be attractive or repulsive, bringing some residues close in space, while keeping others far apart. Consequently, rather than maintaining their inherent flexibility and existing as one-dimensional strings of amino acids, proteins can undergo spontaneous folding into three-dimensional structures.

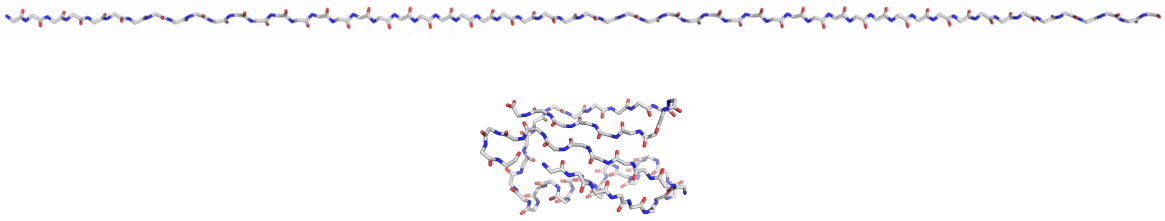


Figure 1.3 Proteins can Fold From a 1-Dimensional Amino Acid Sequence into a 3-Dimensional Structure.
The amino acid sequence for the B1 domain of Protein G (GB1) is shown in both a linear, unfolded and folded form. Structures were generated using PyMOL (version 1.7.2.1, Schrödinger), with the folded structure based on PDB code: 3GB1. For clarity, only the backbone carbon (grey), nitrogen (blue) and oxygen (red) are shown.

Spontaneous, reversible folding of the amino acid chain indicates that the folded state must be thermodynamically favoured over the unfolded state. Analysis of multiple proteins reveals that this difference in Gibbs free energy typically ranges from approximately 2 – 10 kcal mol^{-1} (Maxwell *et al.*, 2005). To put this into context, a single hydrogen bond contributes an energy of around 1 – 5 kcal mol^{-1} (Fersht *et al.*, 1985). Proteins are, therefore, only marginally stable and all forces, even those with small energy contributions, can have a significant influence on protein stability.

1.3.1 Hydrogen Bonding

Hydrogen bonds form between a donor, hydrogen, and an electronegative acceptor group. Each amino acid has a backbone hydrogen bond acceptor, the carbonyl group, and all but proline have a donor, amide N-H group. Further hydrogen bonding capabilities can also be provided in a side chain specific manner. At first glance, it would appear that the ability of the amino acids to form intramolecular hydrogen bonds should produce a highly stable protein. However, the solvent also has to be taken into account (Dill, 1990; Hunter, 2004). For proteins, this is predominantly water (H_2O). By either donating its hydrogen or accepting through its oxygen, water is able to satisfy the amino acid hydrogen bonds. As the bonds can be satisfied by either water or another amino acid, there is little change in free energy upon formation of intramolecular amino acid hydrogen bonds. While hydrogen bonds are therefore not the principal determinant of protein stability, they do provide some distinct advantages for the formation of protein structure. Firstly, hydrogen bonds can be formed between amino acids in a manner that produces a specific structure. This both allows specificity in the conformation of the protein fold, and contributes to the second benefit of intra-protein hydrogen bonds: they can form cooperatively. Establishing a single hydrogen bond can bring other donor and acceptor groups closer in space, promoting the formation of subsequent intra-protein hydrogen bonds.

1.3.2 van der Waals

Orbiting electrons can shift towards one side of an atom, setting up a transient charge. Nearby atoms may be either attracted to, or repulsed from, this charge, creating an electrostatic force known as van der Waals. The strength of van der Waals interactions is related to the distance between interacting atoms (Roth *et al.*,

1996). As 2 atoms become closer in space, the influence on each other's electron orbit increases, causing an increase in electron density shift and a rise in the van der Waals force. This force can be attractive, but once the distance between the two atoms becomes so small that the electron orbits start to overlap, the force between the atoms becomes repulsive. When a protein folds, the interior of the protein typically becomes tightly packed. The reduced distances between atoms results in an enhanced van der Waals interaction force. Consequently, van der Waals forces promote protein folding (Pace *et al.*, 2014).

1.3.3 Electrostatics

Ionisable side chains are present in a subset of amino acids (Figure 1.1). Depending upon the pH, these ionisable groups either gain or lose a proton. Interactions between these residues and other charged groups can be either attractive (opposite charge) or repulsive (same charge). Compared with hydrogen bonds and van der Waals forces, which are both also electrostatic in nature, the force between charged groups functions over relatively large distances. Bringing together oppositely charged residues would therefore be expected to contribute a large driving force for folding. However, the solvent again needs to be taken into account. Oxygen's electronegativity creates a dipole, with oxygen having a net negative charge and hydrogen a net positive charge. In addition to being responsible for the ability of water to hydrogen bond (Chapter 1.3.1), the dipole also allows water to interact with charged amino acid groups. Consequently, electrostatic interactions may not offer a significant driving force for folding, as there is little change in free energy between forming charge-charge interactions with water or within the protein chain. However, electrostatic forces can play a role in setting up the correct protein structure. Burial of either an unsatisfied hydrogen bond or an unpaired charge group in the core of the protein fold comes with a large energy penalty: in the unfolded state the charge/hydrogen bond can interact with water, whereas in the folded state, the bond is broken and the free energy contribution is lost. An unpaired charge is therefore likely to be found on the surface the protein, where it can interact with water. Through this interaction, charged amino acids on solvent exposed surfaces provide the added benefit of improving the protein's solubility.

1.3.4 The Hydrophobic Effect

Compared to the other forces mentioned above, the hydrophobic effect is the main driving force for protein folding (Dill, 1990; Pace *et al.*, 2014). Many of the amino acid side chains are hydrophobic, containing only carbons and hydrogens (Figure 1.1), and are unable to form hydrogen bonds with solvent. When the hydrophobic groups are solvent exposed, this creates an energetic penalty. The source of this penalty is the subject of much debate (Ball, 2011). One hypothesis, supported by recent experiments (Camilloni *et al.*, 2016), states this penalty is a result of reduced solvent entropy due to water molecules maximising their hydrogen bonding by ordering themselves around the hydrophobic residue. The second hypothesis suggests that the hydrophobic groups disrupt the solvent's hydrogen bonding network (Lum *et al.*, 1999), creating an enthalpic penalty. Either way, packing hydrophobic side chains into the protein core removes the interaction with the solvent and the corresponding energetic penalty. Burial of hydrophobic residues is thus associated with a significant gain in free energy, promoting the 3-dimensional folded version of the protein over the 1-dimensional chain (Figure 1.3).

1.3.5 Disulphide Bonds

Containing a thiol group, the side chains of 2 cysteine residues can couple, forming a covalent, disulphide bond. To form this interaction, the 2 cysteines need to be close in space. Consequently, it is perhaps more likely for a native disulphide bond to form when the protein is folded, as residues that may be far apart in the unfolded state are brought together. It is easy to assume that the formation of this bond once the protein folds provides a significant enthalpy gain. However, if the protein unfolds after bond formation, the disulphide is still formed and there is no significant difference in enthalpy between the two states. Instead, the stabilising influence of disulphide bonds is due to the entropic effect on the unfolded state. The bringing together of two residues which were far apart in the unfolded protein carries an entropic cost. This cost is already paid in the folded state, as the 2 cysteine residues are brought close in space by other interactions that contribute to the folding of the protein. Thus, disulphide bond formation has an entropic cost to the unfolded state, not the folded state, resulting in a net stabilisation of the folded protein. There is a caveat - not all disulphide bonds stabilise the folded protein. Disulphide formation can create strain in the folded structure, disrupting other native interactions, or can stabilise the unfolded state by increasing

intramolecular interactions (Clarke *et al.*, 2000). Although some of the proteins investigated in this thesis contain cysteine residues, none of the proteins contained native disulphide bonds.

1.4 Protein Structure

Proteins can experience various levels of structure. The most basic structure is one that all proteins have – the covalent linking of amino acids in a 1-dimensional chain. Known as the primary structure, this is effectively the sequence of amino acids and is traditionally referenced from the amino acid with the free amino group (N-terminus). Whether a protein forms further levels of structure is dependent on the influence the exerted forces have on the rotatable backbone phi and psi angles. However, not all torsion angles are equal. This inequality is superbly demonstrated by the Ramachandran Plot (Figure 1.4), which illustrates the favoured phi and psi angles, and the regions that are inaccessible due to steric clashes (Ramachandran *et al.*, 1963).

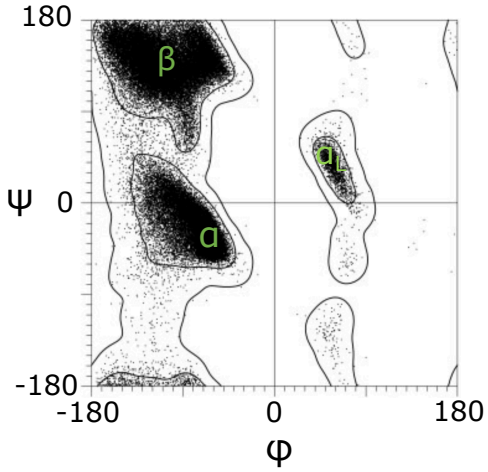


Figure 1.4 Ramachandran Plot Demonstrating the φ and ψ Angles for a Typical Amino Acid.
Due to steric clashes, the phi (φ) and psi (ψ) torsion angles for a typical amino acid only sample a subset of conformations (lines). Analysis of protein structures indicates that amino acids (black points) typically fall within distinct favoured regions (interior lines). These regions represent different secondary structures: alpha-helix (α); beta-sheet (β); and left handed alpha-helix (α_L). The plot was adapted from Lovell *et al* (Lovell *et al.*, 2003) and indicates the allowed and observed phi and psi angles for all amino acids, except proline and glycine.

Except proline, all amino acids contain a hydrogen bond donor, the amide hydrogen, and an acceptor, the carbonyl group. Given the restrictions on the phi and psi angles, intramolecular hydrogen bonds are typically satisfied in one of two ways. Firstly, the carboxyl of one amino acid can interact with the amide hydrogen that is 3 – 4 residues ahead in sequence (Figure 1.5). This creates a helical structure, with a full turn occurring either every 3 residues (3_{10}-helix), or every 3.6 residues (α-helix). Hydrogen bonds can also be satisfied by interacting with residues that are further away in sequence. For example, if a turn is introduced in the amino acid chain, then the amide hydrogen of one residue can interact with the carboxyl of another. This type of structure is named β-sheet and is relatively more extended than the 3_{10}- or α-helix. Due to the differing nature of their hydrogen bond patterns, both helices and β-sheets exhibit distinct phi and psi angles (Figure 1.4). Compared to the random coil-like primary structure, these torsion angles are relatively fixed. Thus, being more ordered versions of the primary structure, both β-sheets and helices represent the second level of protein structure (secondary structure).

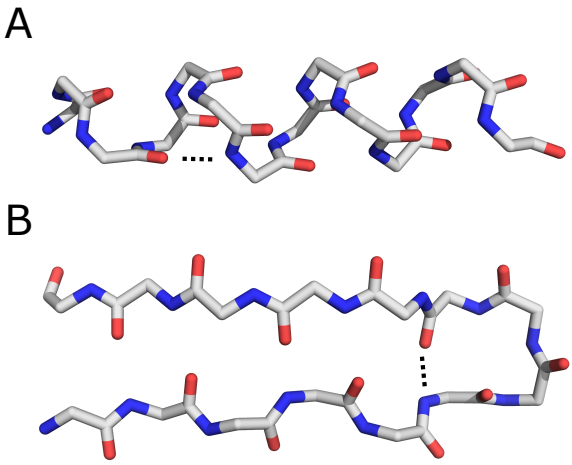


Figure 1.5 Protein Secondary Structure.
Restrictions on torsion angles and hydrogen bonding patterns cause amino acids to be structured in 1 of 2 ways. In an α-helix (A), the amide hydrogens bond with the carboxyl groups that is 4 residues towards the N-terminus. For β-sheets (B), a turn in the polypeptide chain allows hydrogen bonding between the carboxyl groups and amide hydrogens of residues that are relatively separated in sequence. To aid the eye, a single hydrogen bond is indicated with a dashed line, but in both cases, the alignment of the chain allows multiple hydrogen bonds to form. For clarity, only the

backbone carbon (grey), nitrogen (blue) and oxygen (red) atoms are shown. Structures were generated using PyMOL (version 1.7.2.1, Schrödinger).

When the secondary structure elements interact, protein tertiary structure is formed (Figure 1.6). This folding is the highest level of structure achieved by monomeric proteins and is driven by interactions between the amino acid side chains (see Chapter 1.3 for a description of the forces involved).

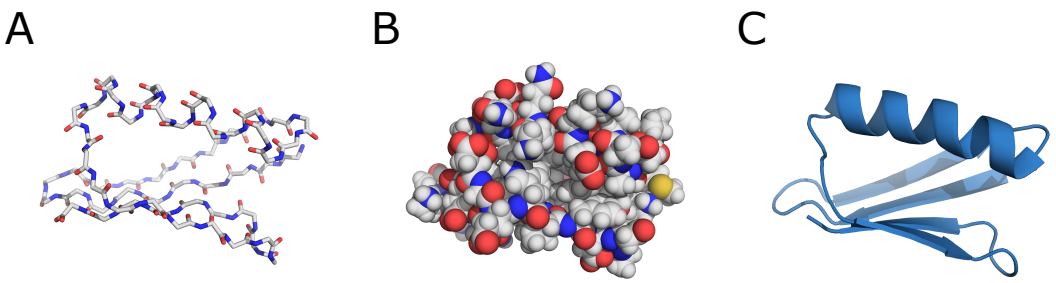


Figure 1.6 Visual Representation of Protein Tertiary Structure.
(A)When GB1 folds, the secondary structural elements of a single α-helix and a four strand β-sheet come together. (B) Showing all atoms indicates that the folded structure is tightly packed, with the side chains of each secondary structure element interacting. (C) To ease comparison between different protein folds, structures can be presented in a cartoon manner, showing only the arrangement of the secondary structure elements. Structures were produced using PyMOL (version 1.7.2.1, Schrödinger) and are based on PDB code: 3GB1.

1.5 Structure-Function Paradigm

The human genome encodes more than 20,000 genes (International Human Genome Sequencing Consortium, 2004). Proteins from almost half of these are expressed in every cell (Uhlen *et al.*, 2015), with approximately 1 million proteins present in every fL (Milo, 2013). A key question arises from this observation – how do proteins carry out specific functions and avoid non-specific interactions with each other?

When a protein folds, it displays a 3-dimensional surface to its environment. Through positioning specific amino acids in specific places on the surface, proteins are able to present a unique interface. For example, residue side chains in the active site of enzymes are typically highly conserved (Bartlett *et al.*, 2002), creating a specific interaction surface. Analogous to a lock and key (Fischer, 1894), this

specificity ensures that only the correct substrate (the key) can fit into and be catalysed by the enzyme's active site (the lock). Problems arise when proteins lose their unique 3-dimensional surface – misfolded or unfolded proteins can stick together forming toxic aggregates or amyloids (Dobson, 2003). Thus, through enabling the precise 3-dimensional localisation of specific amino acids, tertiary structure allows proteins to function specifically, minimise non-specific interactions and reduce the potential for aggregation.

The link between structure and function led to the development of the structure-function paradigm, which states that in order to function, a protein must have tertiary structure. Given the acclaimed importance for protein function, studies of how and why proteins fold were of critical importance.

1.6 Studies of Protein Folding

From a purely statistical standpoint, it seems incredible that a protein could fold into a specific structure simply by chance. As demonstrated by the Ramachandran plot (Figure 1.4), each amino acid can occupy a variety of states. Considering a protein of 101 amino acids, each with just three possible configurations, this provides a total of 5×10^{47} (3^{100}) different combinations. Thus, as stated by Levinthal, even if the rate of exchange between these configurations occurred at a rate of 10^{13} s^{-1}, it would take 10^{27} years to sample all possible states (Zwanzig *et al.*, 1992). Given that this is longer than the age of the universe, protein folding clearly cannot be due to a purely random sampling of all possible bond angles. Instead, proteins are funnelled to their final conformation (Onuchic and Wolynes, 2004). Imagine two amino acids that are distant in the unfolded chain, but close in space in the folded structure. If these two amino acids interact in the unfolded state, due to Brownian motion, they are more likely to have a lower free energy than the interaction of two residues that do not form native interactions. This interaction is therefore more likely to persist and allow other native-like interactions to form (Dobson, 2003). Hence, rather than sampling all possible conformations, the amino acid chain only needs to sample enough to form stabilising interactions, ultimately providing directionality towards the native state.

1.6.1 Phi-Value Analysis

Anfinsen's experiment demonstrated that all of the information required to correctly fold a protein was contained within the amino acid chain (Anfinsen *et al.*, 1961). Given that proteins fold by forming contacts between residues, another question is provoked: are all amino acids equally involved in the folding process? If not, which ones are the crucial, fold encoding residues? The involvement of residues in either the unfolded or folded state can be easily addressed by equilibrium structural studies. However, probing the folding process is a little more problematic. One technique, named phi-value (φ-value) analysis, is able to determine amino acid interactions in the relatively short lived, reaction transition state (Matouschek *et al.*, 1989). The transition state is the point of highest free energy on the reaction coordinate and, for a 2-state process, reaching the transition state is the rate limiting step for either folding or unfolding. To gain insight into the transition state, φ-value analysis utilises mutagenesis, kinetic and equilibrium experiments. For example, imagine a leucine residue that forms all its native-like interactions in the transition state. If the leucine side chain was shortened by mutation (e.g. to alanine), the side chain interactions in both the transition and folded state would be lost. This would destabilise the folded protein and the transition state by the same amount. On the other hand, if the leucine was only making interactions in the folded state, mutation to alanine would only destabilise the folded protein, not the transition state. There is also a scenario between these extremes – the leucine can make some, but not all, of its native-like interactions in the transition state. Or the leucine could make all of its native interactions, but with a lower energy than in the folded state. In these cases, mutation to alanine would destabilise the transition state to a lesser amount than the folded state (Figure 1.7).

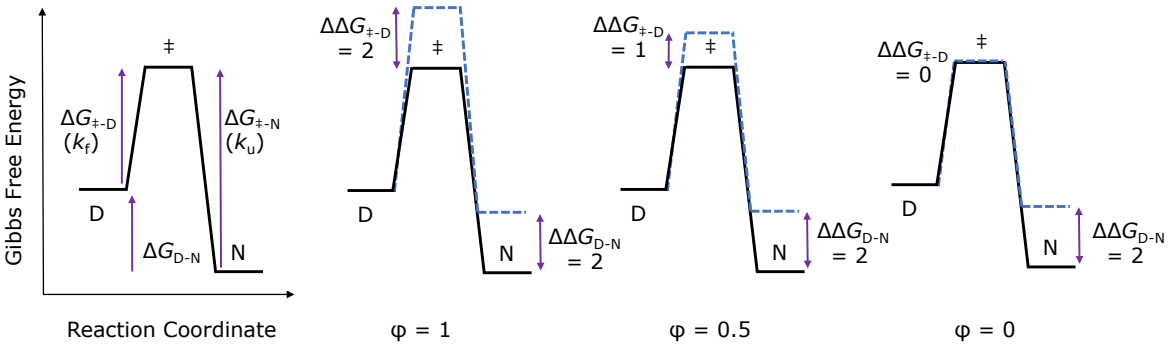


Figure 1.7 Free Energy Diagrams for the Folding of a Protein and the Possible Influence of Mutation.
For a protein that is predominantly folded at equilibrium, the folded (native; N) state of the protein is lower in energy than the unfolded (denatured; D) state. The difference between the two ($\Delta G_{D\text{-}N}$) can be measured by equilibrium chemical denaturation studies and represents the free energy of unfolding. The rate of protein folding (k_f) and unfolding (k_u) are related to the height of the energy barrier between D and the transition state ($\Delta G_{‡\text{-}D}$), and N and the transition state ($\Delta G_{‡\text{-}N}$), respectively. The folding rates can be measured through kinetic chemical denaturation experiments. Examples for the possible effects of mutation (blue, dashed lines) are shown when the residue makes all its native interactions in the transition state (2nd from the left, φ =1); some native interactions in the transition state (2nd from the right, φ = 0.5); or no native interactions in the transition state (right, φ = 0). $\Delta\Delta G$ values are added for illustrative purposes. A description of the calculation for $\Delta\Delta G$ and φ-values is given below.

The $\Delta\Delta G$ values for the mutation are calculated as follows (Equations 1.1, 1.2, 1.3):

$$\Delta\Delta G_{\mathrm{D-N}} = \Delta G_{\mathrm{D-N}}{}^{\mathrm{WT}} - \Delta G_{\mathrm{D-N}}{}^{\mathrm{Mut}} \qquad (1.1)$$

where $\Delta G_{D\text{-}N}{}^{WT}$ is the ΔG of unfolding for wild-type and $\Delta G_{D\text{-}N}{}^{Mut}$ is the ΔG of unfolding for the mutant protein

$$\Delta\Delta G_{\mathrm{D}-\ddagger} = RT\ln\frac{{k_{\mathrm{f}}}^{\mathrm{WT}}}{{k_{\mathrm{f}}}^{\mathrm{Mut}}} \quad (1.2)$$

where R is the gas constant, T is the temperature in kelvin, ${k_{\mathrm{f}}}^{\mathrm{WT}}$ is the folding rate of the wild-type protein and ${k_{\mathrm{f}}}^{\mathrm{Mut}}$ is the folding rate of the mutant protein

$$\Delta\Delta G_{\ddagger-\mathrm{N}} = -RT\ln\frac{{k_{\mathrm{u}}}^{\mathrm{WT}}}{{k_{\mathrm{u}}}^{\mathrm{Mut}}} \quad (1.3)$$

where R is the gas constant, T is the temperature in kelvin, ${k_{\mathrm{u}}}^{\mathrm{WT}}$ is the unfolding rate of the wild-type protein and ${k_{\mathrm{u}}}^{\mathrm{Mut}}$ is the unfolding rate of the mutant protein.

Normalisation of the changes in ΔΔG provides a φ-value (Equation 1.4):

$$\varphi = \frac{\Delta\Delta G_{\mathrm{D}-\ddagger}}{\Delta\Delta G_{\mathrm{D}-\mathrm{N}}} \ \mathrm{Or}\ \varphi = 1 - \frac{\Delta\Delta G_{\ddagger-\mathrm{N}}}{\Delta\Delta G_{\mathrm{D}-\mathrm{N}}} \quad (1.4)$$

Thus, the φ-value provides information on how the interactions of the mutated residue compare in the transition state and the folded (native) state. If the mutated residue makes all of its native interactions in the transition state, a φ-value of 1 is achieved. If no native interactions are formed then a φ-value of 0 is observed. A φ-value between 0 and 1 can also occur when the residue makes some of the native interactions in the transition state, or all of its native interactions, but with a lower energy than in the folded state.

A picture of the transition state can therefore be built up by producing multiple φ-values. Allowing the residues that encode the folding process to be identified.

1.6.1.1 Assumptions of Phi-Value Analysis

When performing or interpreting a phi-value analysis, it is important to remember the basic assumptions of the technique (Fersht *et al.*, 1992). (1) The mutations should not introduce new interactions in either the transition state or the folded state. Otherwise, the calculated $\Delta\Delta G$ values would be a composite of the destabilising effects, due to the loss of the WT residue interaction, and the new interactions introduced by the mutation. (2) Mutation does not alter the pathway of folding. If it did, the reported phi-value would not be representative of the WT transition state. (3) Neither the folded or unfolded state structure should be significantly altered by the mutation. (4) Through taking the ratio of the WT and mutant folding rates in Equations 1.1, 1.2 and 1.3, the pre-exponential factor of the Arrhenius equation can be cancelled. The implicit assumption, therefore, is that the pre-exponential factor is not altered by mutation.

Assumption 1 can be tested by performing multiple mutations. For example, the phi values obtained from mutating an isoleucine to a valine can be compared to the phi-values obtained when mutating the same isoleucine to an alanine. If the valine introduced new interactions, these would be expected to be different or lost in the shorter side chain containing alanine. Consequently, the ratio of the $\Delta\Delta G$ values and the determined phi-value would differ. Alternatively, if no new interactions were introduced by the mutation, both valine and alanine would provide the same phi-value. The second assumption can also be tested by this method or by mutation of an additional, distinct residue. If the additional mutation shifts the folding pathway, the determined phi-value will differ. Whereas, an identical phi-value should be obtained when in the presence of the WT or mutated background, if the pathway is not altered by mutation. Assumption 3 can be assessed using equilibrium structural techniques, such as NMR.

1.6.1.2 Caveats of Phi-Value Analysis

Through providing the ability to explore protein transition states, phi-value analysis has proved a key method for studying protein folding. However, its use has not been without controversy (Cho and Raleigh, 2006; de los Rios *et al.*, 2006; Gianni and Jemth, 2014; Naganathan and Muñoz, 2010; Ozkan *et al.*, 2001; Sánchez and Kiefhaber, 2003). Much of the debate has surrounded the sensitivity and accuracy of phi values. Due to the use of ratios, it has been argued that small errors in either

the numerator or denominator can lead to large errors in phi. This is particularly problematic when the mutation provides little destabilisation. Phi-value analysis, therefore, should be pursued with a few caveats. Firstly, the mutation should provide a $\Delta\Delta G_{D\text{-}N}$ of >0.6 kcal mol^{-1}. Secondly, to prevent large scale changes in transition state structure, conservative mutations should be utilised. Finally, given the potential for error in an individual calculation, phi-values should be interpreted as groups of values at multiple positions, rather than interpreting a single value at a single position. When these caveats are employed, the method can generate reliable transition state information, providing insight into protein folding pathways and mechanisms (Fersht and Sato, 2004; Gianni and Jemth, 2014).

1.7 Intrinsically Disordered Proteins

A consequence of the structure function paradigm is that to fully understand a protein's function, a detailed knowledge of its structure is required. Since 1958, when the first protein structure was published (Kendrew *et al.*, 1960, 1958), over 120,000 protein structures have been deposited in the Protein Data Bank (PDB). Along the way, some proteins were found that lacked structure, but these were generally considered experimental artefacts or rare exceptions.

At the end of the 20th century it started to become clear that unstructured proteins were not irrelevant (Wright and Dyson, 1999). Rather than folding to well-defined three-dimensional structures in physiological conditions, unstructured proteins can remain unstructured in the crowded environment of the cell (Binolfi *et al.*, 2012; Theillet *et al.*, 2016). Given that the fold of a protein is encoded within its amino acid sequence (Anfinsen, 1973), the sequence of unfolded proteins must encode their structure-lacking nature. Compared to folded proteins, unfolded proteins typically contain an excess of proline and charged residues (Romero *et al.*, 2001; Theillet *et al.*, 2014; Uversky *et al.*, 2000). Combined with the relative lack of bulky hydrophobic residues, both the secondary structure propensity and the hydrophobic driving force for folding are reduced. Identification of sequence bias in unfolded proteins indicates that they are not merely unstable folded proteins, but proteins that are encoded to lack structure. In line with this idea, they have been termed intrinsically disordered proteins (IDPs) (Dunker *et al.*, 2001; Schlessinger *et al.*, 2011).

Capitalising on the sequence bias of IDPs, bioinformatic analysis sought to ascertain the prominence of disorder in biology. Remarkably, predictors indicated that disordered proteins were not just rare exceptions of the structure function paradigm. Over 30% of eukaryotic proteins are expected to contain intrinsically disordered regions (IDRs) of >30 amino acids in length (Peng *et al.*, 2013; Ward *et al.*, 2004). Perhaps unsurprisingly, given the 'lock and key hypothesis' (Fischer, 1894), enzymes typically lack disordered regions. Instead, protein disorder is found predominantly in signalling and transcription processes (Ward *et al.*, 2004).

The ability of populations to adapt to their environment is a crucial benefit of protein evolution. However, whilst evolution can provide increased fitness, mutations in protein sequences can also be detrimental. In folded proteins, mutations in the core can have large effects on protein stability (Tokuriki *et al.*, 2007), and can ultimately result in disease (Yue *et al.*, 2005). Consequently, there is a strong evolutionary pressure to maintain the position and identity of amino acids that contribute to the thermodynamic structural stability (Mirny and Shakhnovich, 1999). Disordered proteins are not under the same evolutionary pressure as, by definition, they do not have a fold. This allows IDPs to tolerate more mutations, making them more amenable to evolution (Chen *et al.*, 2006). Yet, despite the relatively rapid rate of mutation, the disordered nature of IDP sequences is conserved (Chen *et al.*, 2006), indicating important functional benefits to disorder beyond the ability to tolerate mutation.

Lacking structure, amino acid side chains in IDPs are typically more accessible than in folded proteins, making them more amenable to post-translational modifications (Bah and Forman-Kay, 2016). This can have a crucial impact on function. For example, the IDP 4E-BP2 suppresses cap-dependent mRNA translation by binding to the eukaryotic translation initiation factor 4E (eIF4E). Phosphorylation of the 4E-BP2 induces folding and sequestering of the interaction motif, reducing binding affinity and promoting translation (Bah *et al.*, 2015).

The ability to undergo weak, but highly specific interactions is another touted functional advantage of disorder (Chen, 2012; Zhou, 2012). Fast binding and unbinding rates could be of particular functional use in transcriptional processes, where the ability to respond quickly to changes in conditions is important (Shammas, 2017). However, while complexes involving disordered proteins bind

less tightly on average, the span of affinities for complexes involving disordered or purely folded proteins is similar (Shammas *et al.*, 2012). Additionally, while the association of complexes involving disordered proteins can be remarkably fast (Shammas *et al.*, 2013), rapid binding and unbinding rates can also be achieved by folded proteins (Schreiber and Fersht, 1996; Shammas *et al.*, 2012).

Due to their roles in important cellular processes, IDPs require tight regulatory mechanisms (Babu *et al.*, 2011). Compared to their folded counterparts, disordered proteins typically tend to have an increased number of ubiquitination sites (Edwards *et al.*, 2009) and higher rates of protein degradation (Tompa *et al.*, 2008). Combined with an enhanced rate of mRNA degradation, this maintains low cellular concentration of IDPs (Edwards *et al.*, 2009). Problems occur when these control mechanisms fail: aberrant expression levels and half-lives of IDPs are associated with multiple diseases, including neurodegeneration and cancer (Uversky *et al.*, 2008). However, as with every generality, it should be noted that this is not the case for all IDPs. Some IDPs are expressed at high levels and perform their function without causing issue (Edwards *et al.*, 2009).

1.7.1 Coupled Folding and Binding

An essential aspect of signalling and transcription processes is the ability of proteins to interact with other molecules. Disordered regions are over-represented in these processes (Ward *et al.*, 2004), suggesting that they may convey specific advantages over folded proteins. Upon interacting with a partner macromolecule, a subset of IDPs fold to a well-defined 3-dimensional structure (Dyson and Wright, 2002). A potential advantage for IDPs over folded proteins is structural plasticity, which allows the same sequence to form different conformations when bound. For example, the disordered C-terminal domain of p53 can bind as a strand (Avalos *et al.*, 2002), a helix (Rustandi *et al.*, 2000) or a coil (Lowe *et al.*, 2002; Mujtaba *et al.*, 2004) (Figure 1.8).

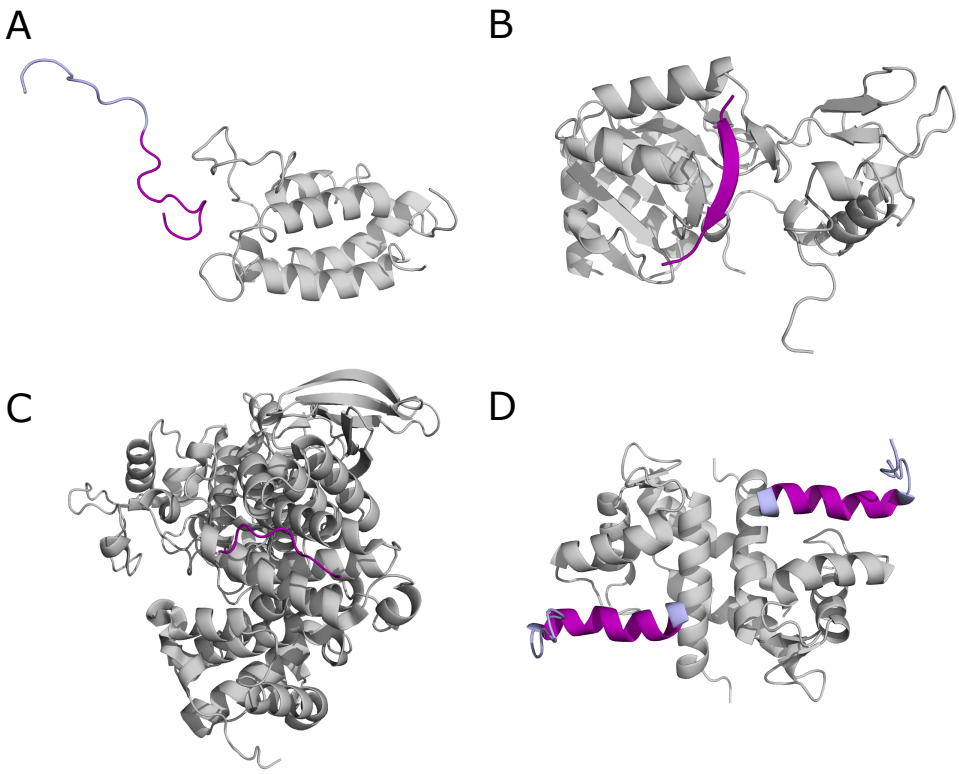


Figure 1.8 p53 Folds and Binds to Many Different Conformations.
Structure of the disordered C-terminal domain of p53 when bound to (A) CBP Bromodomain (PDB code: 1JSP); (B) Sir2 (PDB code: 1MA3); (C) CyclinA (PDB code: 1H26); and (D) S100B(ββ) (PDB code:1DT7). Residues of p53 that are present in all structures are shown in purple, additional residues are shown in lavender. The binding partners are indicated in grey. Structures were produced using PyMOL (version 1.7.2.1, Schrödinger).

IDPs can sample elements of secondary structure, existing as ensembles of multiple conformations. Within the ensemble, the disordered protein may sample a structure that resembles the bound state. Mechanistically, this raises a pertinent question. Is the structure required and formed before binding? Or does the IDP bind and then fold? These 2 scenarios are referred to as conformational selection, and induced fit, respectively (Figure 1.9).

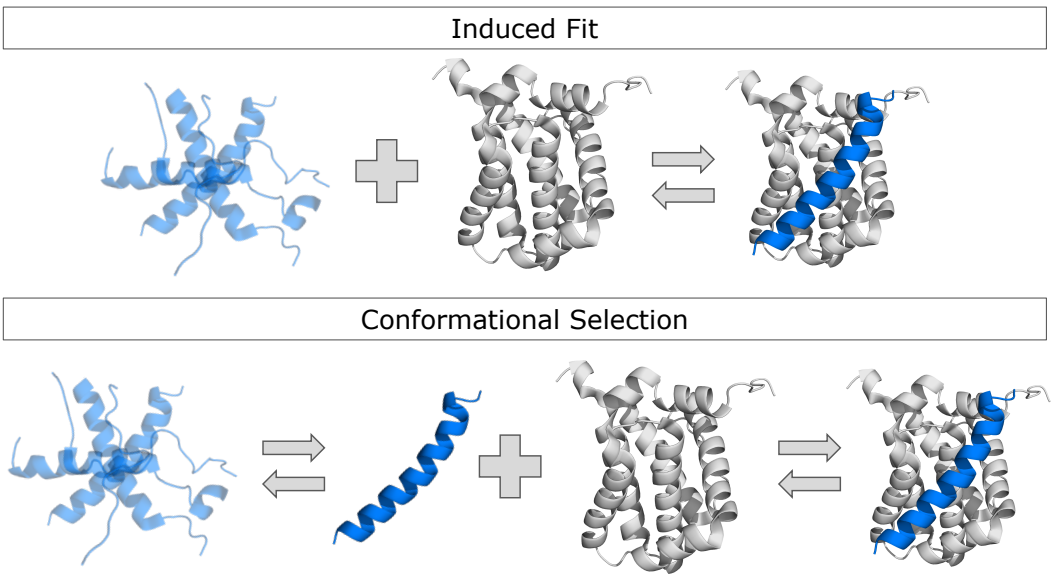


Figure 1.9 Potential Mechanisms for Coupled Folding and Binding.
There are 2 conceivable extremes for IDPs that fold and bind. Firstly, in the induced fit mechanism (top), any conformation within the IDP ensemble is able to bind to its partner and subsequently fold – IDP folding occurs after binding. Secondly, in the conformational selection mechanism (bottom), some IDP molecules within the ensemble may sample a structure that resembles the bound state. Only these molecules are able to bind to their partner – IDP folding occurs before binding. Between these two extremes, a third mechanism is also possible. Binding to the partner could require the IDP to contain some structure. After binding in a partially folded conformation, the IDP would then fold to the final bound structure. Thus, resembling a mixed conformational selection and induced fit mechanism. Structures were produced using PyMOL (version 1.7.2.1, Schrödinger).

1.8 Aims of this Thesis

At the time that the work described in this thesis began (2014), the prevalence of IDPs in biology had already been recognised for more than a decade (Ward *et al.*, 2004; Wright and Dyson, 1999). Disorder brings with it a few issues - the exposed nature of IDPs can promote aggregation and aberrant regulation can result in disease (Babu *et al.*, 2011). Given these complications, the prominence of IDPs and IDRs in signalling and transcription processes indicates that they may have significant advantages over folded proteins. Several benefits had been proposed (Chen, 2012); however, few had been experimentally determined. Debates surrounded the relevance of 'bound-like' structure within IDPs that could undergo coupled folding and binding. Highly cited work claimed that pre-formed structure gave a functional advantage (Fuxreiter *et al.*, 2004), but, if so, why have disorder at all?

This thesis aims to answer some of the outstanding questions. Specifically, (1) how does altering the level of residual 'bound-like' structure alter coupled folding and binding reactions? (2) Are non-contacting residues just bystanders, joining together the interacting residues? Or do they influence coupled folding and binding reactions? (3) Finally, how do reactions involving the folding and binding of an IDP compare to the unimolecular folding of a protein? Where is the folding information encoded? These questions were addressed using equilibrium, kinetic and structural methods.

Chapter 2

Kinetics and Thermodynamics of Folding and Binding

This thesis examines the population of molecules (proteins) in different states under a given set of conditions, and how those populations change when the conditions are altered. Thermodynamics provides a mathematical framework to predict whether a reaction, such as a protein changing state from unfolded to folded, will occur.

2.1 Entropy, Enthalpy and Gibbs Free Energy

As described by the second law of thermodynamics, a reaction will occur spontaneously if it results in an increase in the entropy of the universe. This does not mean that the reaction itself must be associated with an increase in entropy. Instead, the reaction could lead to an increase in the entropy of the surroundings, as shown in Equation 2.1:

$$\Delta S_{\mathrm{uni}} = \Delta S_{\mathrm{sys}} + \Delta S_{\mathrm{sur}} \tag{2.1}$$

where ΔS_{uni} is the change in entropy of the universe, ΔS_{sys} is the change in entropy of the system where the reaction is taking place (e.g. the protein solution), and ΔS_{sur} is the change in entropy of everything else.

The change in entropy of the surroundings can be calculated by dividing the energy change of the reaction by the temperature, as shown in Equation 2.2:

$$\Delta S_{\mathrm{sur}} = \frac{\Delta q_{\mathrm{sur}}}{T} \tag{2.2}$$

where T is the temperature in kelvin and Δq_{sur} is the change in the energy of the system that is transferred as heat to the surroundings (i.e. the energy transferred as heat from the protein solution to the surroundings as the protein folds).

As a protein folds, the nature and number of intramolecular and intermolecular interactions are altered, resulting in a change in the internal energy of the system. Some of this energy may expand or contract the volume of the protein solution, while the rest is transferred as heat. Thus, in a situation where the volume of the protein solution can change upon the protein folding, the total change in the energy of the system is not equal to the energy transferred as heat. As shown in Equation 2.3, by considering both components, enthalpy quantifies the energy transferred as heat to or from the system:

$$\Delta H_{\mathrm{sys}} = \Delta U_{\mathrm{sys}} + p\Delta V_{\mathrm{sys}} \tag{2.3}$$

where ΔH_{sys} is the change in enthalpy of the system, ΔU_{sys} is the change in internal energy of the system, p is the pressure exerted on the system (assumed to be constant) and ΔV_{sys} is the change in volume of the system.

Under reversible conditions, the energy transferred to the solution as heat (ΔH_{sys}) is equal and opposite to the energy transferred as heat from the surroundings (Equation 2.4):

$$\Delta H_{\mathrm{sur}} = -\Delta H_{\mathrm{sys}} \tag{2.4}$$

The negative enthalpy change of the system (i.e. the energy transferred to the surroundings as heat) can therefore be substituted into Equation 2.2, giving Equation 2.5:

$$\Delta S_{\text{sur}} = \frac{-\Delta H_{\text{sys}}}{T} \quad (2.5)$$

Substituting Equation 2.5 into Equation 2.1 gives Equation 2.6:

$$\Delta S_{\text{uni}} = \Delta S_{\text{sys}} + \frac{-\Delta H_{\text{sys}}}{T} \quad (2.6)$$

Multiplying Equation 2.6 through by -*T* gives Equation 2.7.

$$-\text{T}\Delta S_{\text{uni}} = -T\Delta S_{\text{sys}} + \Delta H_{\text{sys}} \quad (2.7)$$

This equation is known as the Gibbs function, or the Gibbs free energy (Equation 2.8), and is named after Josiah Willard Gibbs. As the Gibbs free energy (*G*) relates to the entropy and enthalpy of the system, the subscripts can be dropped.

$$\Delta G = \Delta H - T\Delta S \quad (2.8)$$

Assuming a constant temperature and pressure, the Gibbs free energy relates the change in entropy of the universe to the change in entropy and enthalpy of the system. Thus, any reaction that has a negative change in the Gibbs free energy (ΔG) will result in an increase in the entropy of the universe. Reactions associated with a negative ΔG will therefore take place spontaneously. Enthalpy has units of kilojoule per mole (kJ mol^{-1}), entropy has units of joule per kelvin (JK^{-1}) and Gibbs free energy has units of kJ mol^{-1}. In protein folding studies, joules are typically replaced by calories. The 2 units can be inter-converted using a factor of 1 joule being equal to 4.18 calories.

2.2 Populations at Equilibrium

A reaction will occur spontaneously if it is associated with a negative change in the Gibbs free energy. However, as the reaction proceeds, the number of reactants reduces and the number of products increases. For a reversible reaction such as a protein folding, the reaction proceeds in both the forward (folding) and reverse (unfolding direction), as shown in Equation 2.9:

$$\mathrm{D} \rightleftharpoons \mathrm{N} \tag{2.9}$$

Where D is the unfolded (denatured) state and N is the folded (native) state.

The reaction of unfolded protein (D) forming folded protein (N) proceeds until the forward flux is equal to the reverse flux. At this point, the reaction has reached equilibrium. Taking the concentration ratio of the 2 species at equilibrium provides a new parameter, the equilibrium constant, K_{eq}. For example, for a protein folding, K_{eq} can be calculated using Equation 2.10:

$$K_{\mathrm{eq}} = \frac{[D]}{[N]} \tag{2.10}$$

where $[N]$ is the concentration of native, folded protein and $[D]$ is the concentration of denatured, unfolded protein.

This thesis also considers the folding and binding reactions of IDPs with structured partner proteins. In these cases, the reactions are defined by Equation 2.11. K_{eq} of these reactions can be described either in the forward direction (K_a) (Equation 2.12), or more commonly, in the reverse direction (K_d) (Equation 2.13):

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{C} \tag{2.11}$$

$$K_{\mathrm{a}} = \frac{[C]}{[A][B]} \tag{2.12}$$

$$K_d = \frac{[A][B]}{[C]} \tag{2.13}$$

where [*A*] is the concentration of protein A, [*B*] is the concentration of protein B and [*C*] is the concentration of the complex formed between proteins A and B. The concentrations are in units of molar (M). It is for this reason that the dissociation constant (K_d; units of M) is more convenient and easier to interpret than the association constant (K_a; units of M^{-1}).

As the equilibrium constant is the ratio of products over reactants, it provides information relating to the position of the equilibrium i.e. the amount of product vs the amount of reactant. In the case of the equilibrium dissociation constant (K_d), lower numbers indicate a high affinity complex, where the product (the complex) is favoured over the reactants. Equation 2.13 provides another piece of information: when [*A*] equals K_d, [*B*] is equal to [*C*]. Thus, the K_d is the concentration of A required to form a complex with 50% of the molecules of protein B.

The equilibrium constant for a reaction can be converted into a difference in Gibbs free energy between the products and reactants using Equation 2.14:

$$\Delta G = -RT\ln K \tag{2.14}$$

where *R* is the ideal gas constant (8.31 J K^{-1} mol^{-1}), *T* is the temperature in kelvin and *K* is the equilibrium constant for the reaction.

Hence, as well as indicating whether a reaction is spontaneous, the Gibbs free energy describes the position of the equilibrium: the more negative the value of Δ*G,* the larger the ratio of product(s) *vs* reactant(s).

2.2.1 Measuring *K*

Proteins can contain the naturally fluorescent amino acids, tyrosine and tryptophan. When these fluorescent amino acids change environment, for example, transitioning from solvent exposed into a hydrophobic core during protein folding, the fluorescence intensity and/or the emission wavelength can shift. Consequently, intrinsic fluorescence can be used as a proxy for determining whether the protein is in an unfolded or folded state.

Folded proteins typically have a Gibbs free energy difference of 2 – 10 kcal.mol^{-1} between the unfolded and folded state (Maxwell *et al.*, 2005). Consequently, the concentration of folded protein may be several orders of magnitude greater than the concentration of unfolded protein (~2 x 10^7 excess for 10 kcal.mol^{-1}). This can make it difficult to measure the concentration of denatured protein, and therefore *K* (Equation 2.10), under physiological conditions. To overcome this obstacle, difference in Gibbs free energy between the unfolded and folded protein can be reduced by adding chemical denaturants. As the observed unfolded state is induced by denaturant, it is often referred to as the denatured state. It has been shown empirically that the difference in Gibbs free energy is generally linearly related to the concentration of denaturant (Fersht, 1999), as show in Equation 2.15:

$$\Delta G_{\mathrm{D-N}} = \Delta G_{\mathrm{D-N}}^{\mathrm{H_2O}} - m[\mathrm{Den}] \qquad (2.15)$$

where $\Delta G_{\mathrm{D\text{-}N}}$ is the Gibss free energy difference between the denatured and native state under the experimental conditions, $\Delta G_{\mathrm{D-N}}^{\mathrm{H_2O}}$ is the free energy difference in buffer, *m* is a constant and [Den] is the concentration of denaturant.

Through altering the ratio of unfolded, denatured protein to folded, native protein, the equilibrium constant for that specific experimental condition can be ascertained more accurately. Calculating $\Delta G_{\mathrm{D\text{-}N}}$ at multiple denaturant concentrations allows the

m-value to be determined. The Gibbs free energy difference between denatured and native states in buffer ($\Delta G_{\mathrm{D-N}}^{\mathrm{H_2O}}$) can then be determined by multiplying the *m*-value by the concentration of denaturant required to equally populate both the denatured and native state ($\Delta G_{\mathrm{D\text{-}N}} = 0$).

For bimolecular reactions, measuring *K* can be a little easier. As the formation of the complex requires both proteins, the concentration of the free and bound proteins can be altered by changing the concentration of one of the protein partners.

2.3 Reaction Rates

For a reversible reaction, such as those described in Equations 2.9 and 2.11, the flux of the forward and reverse reactions can be described by reaction rate constants, *k*. For protein folding, these can be termed k_{f}, the rate constant for folding, and k_{u}, the rate constant for unfolding. To distinguish from this terminology, the reaction rates for a 2-state bimolecular reaction are referred to as the association rate constant, k_{on} (the rate of complex formation) and the dissociation rate constant, k_{off} (the rate of complex dissociation).

When a protein can only exist in 2-states, either natively folded or unfolded (denatured), the change in concentration of denatured protein with time is described by Equation 2.16:

$$\frac{\mathrm{d}[D]}{\mathrm{d}t} = k_{\mathrm{u}}[N] - k_{\mathrm{f}}[D] \tag{2.16}$$

where [*N*] is the concentration of natively folded protein, k_{u} is the rate constant for unfolding, [*D*] is the concentration of denatured protein and k_{f} is the rate constant for folding.

At equilibrium, the change in concentration of native and denatured protein is equal. Therefore, as shown in Equations 2.17 and 2.18, the equilibrium constant for a 2-state unimolecular reaction is equal to the ratio of the kinetic rates:

$$0 = k_{\mathrm{u}}[N] - k_{\mathrm{f}}[D] \tag{2.17}$$

$$K = \frac{[D]}{[N]} = \frac{k_{\mathrm{u}}}{k_{\mathrm{f}}} \tag{2.18}$$

When an extra protein molecule is involved, such as in IDP coupled folding and binding reactions, the rate equation becomes a little more complex. Rather than following first order reaction kinetics, the rate of complex formation for a reversible bimolecular reaction is dependent on the concentration of both proteins. Consequently, for the reaction scheme shown in Equation 2.11 – where proteins A and B react to form the complex, C – the change in the concentration of the protein A with time is described by Equation 2.19:

$$\frac{\mathrm{d}[A]}{\mathrm{d}t} = k_{\mathrm{off}}[C] - k_{\mathrm{on}}[A][B] \tag{2.19}$$

where $[C]$ is the concentration of complex, $[A]$ is the concentration of protein A, $[B]$ is the concentration of protein B, k_{on} is the association rate constant for the reaction and k_{off} is the dissociation rate constant for the reaction.

Again, at equilibrium the change in concentration of protein A with time is 0. Therefore, as shown in Equations 2.20 and 2.21, the equilibrium constant is equal to the ratio of the kinetic rates:

$$0 = k_{\mathrm{off}}[C] - k_{\mathrm{on}}[A][B] \tag{2.20}$$

$$K_d = \frac{[A][B]}{[C]} = \frac{k_{\mathrm{off}}}{k_{on}} \tag{2.21}$$

2.3.1 Measuring *k*

When a protein is transferred into conditions that favour the unfolded state, for example, by rapidly mixing the protein with a chemical denaturant, the change in fluorescence with time can be followed. The change in fluorescence is described by an exponential function and can be derived from the rate law, as shown in Equations 2.22, 2.23, 2.24 and 2.25. First, as the reaction is 2-state, the concentration of native protein can be replaced by the concentration of denatured protein at time 0, minus the concentration of denatured protein at time *t*.

$$\frac{\mathrm{d}[D]}{\mathrm{d}t} = k_{\mathrm{u}}([D]_0 - [D]) - k_{\mathrm{f}}[D] \tag{2.22}$$

The equation can then be rearranged to give;

$$\frac{\mathrm{d}[D]}{\mathrm{d}t} + (k_{\mathrm{u}} + k_{\mathrm{f}})[D] = k_{\mathrm{u}}[D]_0 \tag{2.23}$$

integration then gives;

$$[D] = \frac{k_u[D]_0}{k_{u+}k_f} + C\mathrm{e}^{(-k_{u+}k_f)t} \tag{2.24}$$

which is equivalent to a single exponential function:

$$[D] = F_{\mathrm{final}} + \Delta F\mathrm{e}^{(-k_{\mathrm{obs}})t} \tag{2.25}$$

where F_{final} is the fluorescence at the end of the reaction, ΔF is the amplitude of the fluorescence change, k_{obs} is the observed rate constant ($k_{\mathrm{u}} + k_{\mathrm{f}}$) and t is time.

As bimolecular reactions are dependent on the concentration of both individual proteins, integration of the rate law becomes a little more complex. For example, Equations 2.26, 2.27, 2.28 and 2.29 show the model used by Shammas *et al.*

(2013) to fit the association kinetic data for a protein (A) binding to a different protein (B) under reversible conditions;

$$F = F_0 + \Delta F \frac{(b-z)(1-\mathrm{e}^{(zk_{\mathrm{on}}t)})}{2\left(\left(\frac{b-z}{b+z}\right)\mathrm{e}^{(zk_{\mathrm{on}}t)}-1\right)} \tag{2.26}$$

where *b* is equal to;

$$b = -(K_{\mathrm{d}} + (1+x)[A]) \tag{2.27}$$

z is equal to;

$$z = \sqrt{{K_{\mathrm{d}}}^2 + 2(1+x)K_d[A] + (x^2 - 2x + 1)[A]^2} \tag{2.28}$$

x is equal to:

$$x = \frac{[B]}{[A]} \tag{2.29}$$

and $[A]$ is the concentration of protein A and $[B]$ is the concentration of protein B.

This model describes data obtained from experiments where the two proteins are mixed at near equimolar concentrations. When the concentration of one protein is in significant excess over the concentration of the other, then the model can be simplified. For example, if protein B is in a 10-fold concentration excess over protein A, then a maximum of 10% of protein B will go into complex (assuming a 1:1 stoichiometric ratio). The concentration of protein B is therefore effectively constant and can be combined into the association rate constant, as shown in Equation 2.30:

$$\frac{\mathrm{d}[A]}{\mathrm{d}t} = k_{\mathrm{off}}[C] - k'_{\mathrm{on}}[A] \tag{2.30}$$

where k'_{on} is equal to k_{on} multiplied by $[B]$.

Equation 2.30 resembles 2.16 and the reaction can be described by an exponential function (Equation 2.25). Maintaining one of the proteins at a ten-fold concentration excess over the other protein therefore results in the reaction following pseudo-first order kinetics.

2.3.2 Determining ΔΔ*G* from kinetics

Reactions rates are related to the difference in free energy between the transition state and the 2 ground states (e.g. folded or denatured protein). A conversion between reaction rates and free energy barriers is provided by the Arrhenius equation, as shown in Equations 2.31 and 2.32:

$$k_{\mathrm{f}} = A\mathrm{e}^{\left(\frac{\Delta G_{\mathrm{D}-\ddagger}}{RT}\right)} \quad \text{or} \quad \Delta G_{\mathrm{D}-\ddagger} = RT(\ln k_{\mathrm{f}} - \ln A) \tag{2.31}$$

$$k_u = A\mathrm{e}^{\left(\frac{-\Delta G_{\ddagger-\mathrm{N}}}{RT}\right)} \quad \text{or} \quad \Delta G_{\ddagger-\mathrm{N}} = -RT(\ln k_{\mathrm{u}} - \ln A) \tag{2.32}$$

where $\Delta G_{\mathrm{D}\text{-}\ddagger}$ is the difference in free energy between the transition state and the denatured state, $\Delta G_{\ddagger\text{-}\mathrm{N}}$ is the difference in free energy between the native state and the transition state, R is the ideal gas constant, T is the temperature in kelvin, k_{f} is the rate constant for folding, k_{u} is the rate constant for unfolding and A is a constant.

The pre-exponential factor A relates to the reaction rate at infinite temperature, i.e. when the exponent tends to 0. This cannot be directly determined from a simple kinetic experiment. However, when calculating $\Delta\Delta G_{‡\text{-N}}$ or $\Delta\Delta G_{\text{D-}‡}$ for a mutant, the pre-exponential factor is assumed to be equal and can be cancelled, giving Equations 1.2 and 1.3, which are reproduced here for clarity:

$$\Delta\Delta G_{\mathrm{D}-‡} = RT\ln\frac{{k_\mathrm{f}}^{\mathrm{WT}}}{{k_\mathrm{f}}^{\mathrm{Mut}}} \tag{1.2}$$

$$\Delta\Delta G_{‡-\mathrm{N}} = -RT\ln\frac{{k_\mathrm{u}}^{\mathrm{WT}}}{{k_\mathrm{u}}^{\mathrm{Mut}}} \tag{1.3}$$

Chapter 3

Materials and Methods

3.1 Buffers and Reagents

All stock buffers were prepared volumetrically, filter sterilised through 0.22 µM cellulose acetate filters (Sartorius Stedim Biotech), and stored at 25°C. Biophysical buffer used in investigating helix-flanking prolines was 100 mM sodium phosphate with 0.05% tween 20 at pH 7.4. All experiments with BCL-2-like or BH3-containing proteins were performed with a biophysical buffer consisting of 50 mM sodium phosphate with 0.05% Tween 20 at pH 7.0.

Ultrapure deionised water: Deionised water (ddH_2O), defined as having a resistivity of 18.2 MΩ•cm at 25°C, was obtained from water purifiers and used in the preparation of all buffers, broths and reagents, unless otherwise stated.

Lysogeny broth (LB): Media containing 10 g/L tryptone, 5 g/L yeast extract and 10 g/L NaCl (all MP Biomedicals) was prepared and autoclaved at 121°C for at least 20 minutes.

2x yeast extract and tryptone broth (2x YT): Media containing 16 g/L Tryptone, 10 g/L yeast extract and 5 g/L NaCl (all MP Biomedicals) was prepared and autoclaved at 121°C for at least 20 minutes.

Imidazole: A 1 M (34.04 g/L) stock was prepared.

Phosphate buffered saline (PBS), pH 7.4: A 20x stock was prepared containing 162 mM (28.82 g/L) disodium hydrogen phosphate dihydrate (Na_2HPO_4 $2H_2O$), 38 mM (5.93 g/L for the dihydrate) sodium dihydrogen phosphate dihydrate (NaH_2PO_4 $2H_2O$), 54 mM (4.03 g/L) potassium chloride (KCl) and 2.73 M (159.54 g/L) sodium chloride (NaCl).

PBS imidazole buffer: 1x PBS with 25 mM or 500mM imidazole was prepared by dilution of the 20x PBS and 1M imidazole stock buffers.

Phosphate buffer, pH 7.4: A 100 mM stock solution was prepared containing 81 mM (14.42 g/L for the dihydrate) disodium hydrogen phosphate dihydrate (Na_2HPO_4 $2H_2O$) and 19 mM (2.96 g/L) sodium dihydrogen phosphate dihydrate (NaH_2PO_4 $2H_2O$).

Biophysical 100 mM phosphate tween buffer, pH 7.4: Tween® 20 (Fisher) was added to the 100 mM phosphate buffer stock to give a final tween concentration of 0.05% (500 µL per L of phosphate buffer).

Biophysical 50 mM phosphate tween buffer, pH 7.0: A 50 mM buffer was prepared containing 29.7 mM disodium hydrogen phosphate dihydrate (Na_2HPO_4 $2H_2O$, 5.29 g/L), 20.3 mM sodium dihydrogen phosphate dihydrate (NaH_2PO_4 $2H_2O$, 3.17 g/L) and 500 µL of Tween® 20 (Fisher) per 1 L of phosphate buffer.

Tris(hydroxymethyl)aminomethane (TRIS) buffer, pH 7.0: A 200 mM, 10x stock, was prepared containing 16 mM (1.94 g/L) TRIS base and 183 mM (28.8 g/L) TRIS hydrochloride.

TRIS ion exchange buffer, pH 7.0: 1x TRIS, by dilution of the 10x TRIS stock, was prepared containing 0 or 500mM (29.22 g/L) or 1M (58.44 g/L) NaCl.

TRIS buffer, pH 8: A 100 mM stock, was prepared containing 39 mM (4.72 g/L) TRIS base and 60 mM (9.46 g/L) TRIS hydrochloride.

4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid ion exchange buffer (HEPES), pH 7.5: a 10 mM buffer containing 5.7 mM (1.36 g/L) of HEPES free acid ($C_8H_{18}N_2O_4S$) and 4.2 mM (1.09 g/L) of HEPES sodium salt ($C_8H_{17}N_2NaO_4S$) was prepared. For the 10 mM HEPES pH 7.5 1 M NaCl buffer, 58.44 g/L of NaCl was added.

Factor Xa cleavage buffer: A 20 mM pH 8.0, 50 mM NaCl, 5 mM $CaCl_2$ buffer was prepared by dilution of the 100 mM TRIS pH 8.0 stock and addition of 2.92 g/L of NaCl and 0.74 g/L of $CaCl_2$.

TEV cleavage buffer: A 10 mM TRIS pH 8.0, 150 mM NaCl buffer was prepared by dilution of the 10x TRIS stock and addition of 8.77 g/L of NaCl.

M9 Salts: A 10x stock was prepared containing 423 mM disodium hydrogen phosphate dihydrate (Na_2HPO_4 $2H_2O$, 75.3 g/L), 172 potassium phosphate dibasic (HK_2O_4P, 30 g/L) and 85.6 mM NaCl (5 g/L). The stock was adjusted to pH 7.4 by titration with NaOH.

M9 Minimal Media: 1 L of minimal growth media was prepared containing 1 x M9 salts (100 mL of 10x M9 salts), 2 mM MgS04 (2 mL of 1 M stock) 0.1 mM CaCl2 (2 mL of 50 mM stock), 10 µM FeCl3 (1 mL of 10 mM stock), 1 µg/mL of Vitamin B1 (400 µL of 5 mg/mL stock), 1 g of ^{15}N ammonium chloride and 2 g of ^{13}C glucose. The pH was adjusted to 7.4 by titration with HCl.

Alexa Fluor® 594 C5 maleimide dye (Life Technologies): An 11 mM stock was prepared in DMSO, flash frozen in liquid Nitrogen (N_2) and stored in 10 µL aliquots at – 80°C.

Ampicillin (Fisher Scientific): A 100 mg/mL stock solution was prepared, filter sterilised through 0.22 µM polyethersulfone membrane (PEM) filters (Millex), and stored in 1 mL aliquots at – 20°C.

Isopropyl β-D-thiogalactoside (IPTG) (Thermo Scientific): A 1M (238.3 g/L) stock was prepared, filter sterilised through 0.22 µM PEM filters (Millex), and stored at -20°C in 400 µL aliquots.

Thrombin (Sigma): A stock of 1 unit/µL was prepared, filter sterilised filter sterilised through 0.22 µM PEM filters (Millex), and stored in 200 µL aliquots at – 20°C.

TEV protease: tobacco etch virus (TEV) protease was expressed and purified in the lab by Tristan Kwan, according to a slightly modified version of an optimised protocol (Blommel and Fox, 2007). TEV was stored at –20°C in 20 mM TRIS pH 7.5, 125 mM NaCl, 2mM DTT, 2 mM EDTA and and 50% (v/v) glycerol.

Dithiothreitol (DTT) (Fisher Scientific): A 1 M stock solution (154.24 g/L) was prepared in biophysical 50 mM phosphate tween buffer, filter sterilised through 0.22 µM polyethersulfone membrane (PEM) filters (Millex), and stored in 0.25 mL aliquots at – 20°C.

3.2 Escherichia Coli Cells

Three strains of *Escherichia coli* (*E. coli*) cells were used in this work. XL1-Blue cells (genotype: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *proAB lacI*^{q}ZΔ*M15* Tn*10* (Tet^{r})]) were used for site directed mutagenesis. These cells were utilised as they are deficient in endonuclease (*endA*), and consequently, produce good quality miniprep DNA. For protein expression and purification, C41 (genotype: F– *ompT gal dcm hsdS*$_B$(r_B-m_B^-) (DE3)) or C41 (DE3) pLysS (F– *ompT gal dcm hsdS*$_B$(r_B-m_B^-) (DE3) pLysS (Cm^{r})) cells were used. C41 and C41 pLysS cells are derived from BL21 (DE3) cells. Compared to BL21 cells, they contain two mutations in the LacUV5 promoter, which helps to improve the expression of toxic proteins (Miroux and Walker, 1996; Rosano and Ceccarelli, 2014).

3.3 Generation of $CaCl_2$ Competent Cells

E. coli cells were incubated, shaking, at 37°C overnight in 5 mL of Lysogeny broth (LB). Under sterile conditions, 1 mL of overnight culture was added to 100 mL of LB. Cells were incubated at 37°C, shaking, until the optical density at 600 nm (OD_{600}) reached 0.3 – 0.5: when the cells were placed on ice for at least 20 minutes. After splitting the volume into two 50 mL conical tubes, the cells were pelleted by centrifugation at 3,000 rpm (1962 x g) for 10 minutes at 4°C. Pellets were re-suspended in 10 mL of 100 mM $CaCl_2$ 15% glycerol (v/v) before leaving on ice for 20 minutes. After pelleting again, the supernatant was discarded and the cells were re-suspended in 2 mL of 100 mM $CaCl_2$ 15% glycerol (v/v). Cells were aliquoted, flash frozen in liquid N_2 and stored at – 80°C.

3.4 Transformation into Competent Cells

Unless stated otherwise, all transformations were carried out as follows. Plasmid Deoxyribonucleic acid (DNA) was mixed with $CaCl_2$ competent *E. coli* cells and left on ice for 10 minutes. Incorporation of DNA was achieved by heat shock at 42°C for 45 seconds, before placing on ice for 2 minutes. Cells were incubated in 900 µL

of 2x YT for 45 minutes, shaking, at 37°C. After pelleting the cells, the supernatant was removed and the cells re-suspended in 50 – 100 µL of 2x YT, before spreading onto a 2x YT agar plate containing 0.1 mg/mL ampicillin and incubating overnight at 37°C.

3.5 Mutagenesis

Mutants were constructed from wild type (WT) plasmid templates using site-directed mutagenesis. Primers were designed to be more than 30 residues in length, begin and end in at least one G or C and contain the nucleotides to be mutated in approximately the middle of the sequence. Mutation reaction mixtures consisted of 2.5 µL of 10 µM forward and reverse primer mix, 5µl of 2mM nucleotides (Fermentas), 1µL of at least 50 ng/µL plasmid DNA template, 5 µL of 10 x Pfu reaction buffer MgCl2 (ThermoScientific), 1 µL of Pfu Turbo polymerase (ThermoScientific) and 35.5 µL of ddH_2O. Mutation reactions consisted of 15 cycles (95°C for 1 minute, 50°C for 1 minute and 72°C for 7 minutes) and were performed using a PTC-100 programmable thermal controller (Bio-Rad). Parental methylated DNA template was digested by incubation with 10 units of Dpn1 for at least 30 minutes at 37°C. Plasmids (10 µL) were transformed into XL1-blue cells (100 µL). Single colonies were selected from 2x YT agar plates and incubated at 37°C overnight in 2x YT broth containing 0.1 mg/mL ampicillin. Plasmid DNA was isolated and purified using a QIAprep Spin Miniprep Kit (Qiagen), following manufacturer's instructions: briefly, cells were lysed, ribonucleic acid (RNA) was degraded, and proteins were denatured. Cell debris was precipitated and soluble plasmid DNA bound onto a silica membrane. After washing, the DNA was eluted in ddH_2O. Isolated plasmid DNA was sent for sequencing (Beckman Coulter) to confirm successful mutagenesis.

3.6 Protein Expression and Purification

Unless otherwise stated in the subsections below, all proteins and peptides were expressed and purified as follows. Plasmid DNA was transformed into C41 (DE3) or C41 (DE3) pLysS *E. coli* cells, as described above. Protein transcription was typically under the control of a T7 promoter: C41 (DE3) and C41 (DE3) pLysS cells are engineered to be able to express T7 polymerase, which binds to the T7 promoter, leading to transcription. Protein expression during cell growth was inhibited by the lac repressor, which binds to the lac operon and prevents T7 polymerase from being

produced. Cells were grown, shaking, at 37°C to an OD_{600} of 0.4 - 0.6 in 2x YT or LB with 100 µg/mL of ampicillin. Protein expression was then induced by addition of IPTG to a final concentration of 0.1 or 1 mM: IPTG binds to the lac repressor, which disrupts its interaction with the lac operon, leading to T7 polymerase production and transcription of plasmid DNA. Cells were harvested by centrifugation at 5,000 revolutions per minute (rpm) for 45 minutes using a SLC4000 rotor (Sorvall). Cell pellets were flash frozen in liquid N_2 and stored at – 80°C. Cells were lysed on ice using an XL-2020 sonicator (Misonix). Sonication consisted of 3 minutes pulsing, which was split into 15 second (s) pulses separated by 45 s rest. Cell sonicate was pelleted by centrifugation at 18,000 rpm for 45 minutes using a SS34 rotor (Thermo Scientific). Sonicate supernatants were filtered through 0.45 and 0.22 µm PEM filters (Millex) before affinity purification: all recombinantly expressed proteins contained a hexahistadine tag sequence (HisTag) or glutathione S-transferase tag sequence (GST Tag) N-terminal to the protein. Protein purity was assessed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). After the first purification of a new protein, a sample was sent for analysis by mass spectrometry (Department of Chemistry, University of Cambridge) to confirm the identity of the purified product. Specific expression and purification procedures for each of the proteins and peptides are listed below.

3.6.1 CBP KIX

The gene for the kinase-inducible domain interacting domain (KIX) of mouse CREB-binding protein (CBP) (Uniprot P45481, residues 586 – 672) had previously been cloned into a modified version of the pRSET A vector, which contained a HisTag N-terminal to CPB KIX, by Dr Sarah Shammas. A single C41 colony was added to 5 mL of 2x YT with 100 µg/mL ampicillin and left shaking at 37°C overnight. Each 5 mL starter culture was used to inoculate 1L of 2x YT. After induction, cells were incubated, shaking, for four hours at 37°C before harvesting. Cells were thawed and re-suspended in PBS 25 mM imidazole before sonication. Sonicate supernatants were loaded onto a 5 mL HisTrap™ HP column (GE Healthcare) using an ÄKTA fast protein liquid chromatography (FPLC) (Amersham Biosciences). CBP KIX was eluted from the column using a linear gradient of PBS imidazole from 25 to 500 mM over 50 mL. Peak elution was observed at approximately 400 mM imidazole. CBP KIX fractions were pooled and incubated with 200 units of thrombin overnight at room temperature. Finally, cleaved KIX was

loaded onto a HiLoad™ 26/600 Superdex™ 75 pg gel filtration column that had been pre-equilibrated in 100 mM sodium phosphate buffer, pH 7.4. Purified KIX protein (Figure 3.1) was stored at 4°C.

GSGVRKGWHEHVTQDLRSHLVHKLVQAIFPTPDPAALKDRRMENLVAYAKKVEGDM
YESANSRDEYYHLLAEKIYKIQKELEEKRRSRL

Figure 3.1 Amino Acid Sequence for Recombinant CBP KIX.
CBP KIX consisted of 89 amino acids: glycine and serine at the N-terminus (black), as remnants of thrombin cleavage, and residues 586 – 672 of CBP (blue) (Uniprot P45481).

3.6.2 c-MYB

The transactivation domain (TAD) of mouse c-MYB (Uniprot P06876, residues 275 – 327) was expressed as a GB1 fusion protein. The fusion protein gene had previously been cloned into a modified version of the pRSET A vector, which contained a HisTag N-terminal to GB1, by Dr Sarah Shammas. A scrape of C41 cells was added to 5 mL of 2x YT with 100 µg/mL ampicillin and left shaking at 37°C for 2 hours. Each 5 mL starter culture was used to inoculate 1 L of 2x YT. After induction with 0.1 mM IPTG, cells were incubated overnight, shaking, at 18°C, before harvesting. Cells were thawed and re-suspended in PBS 25 mM imidazole before sonication. Sonicated supernatants were incubated with 3.5 mL of nickel agarose resin (Agarose Bead Technologies) for at least 1 hour at 4°C. Resin was washed three times with PBS 25mM imidazole and twice with 20mM TRIS pH 7 10mM imidazole, before incubating overnight at room temperature with 100 units of thrombin. Cleaved c-MYB was loaded onto a 5 mL HiTrap™ Q HP ion exchange column (GE healthcare) using an ÄKTA FPLC (Amersham Biosciences). A TRIS ion exchange buffer linear gradient of 0 to 300 mM NaCl over 50 mL was used to elute bound c-MYB. Peak elution was observed at approximately 110 mM NaCl. Finally, c-MYB was loaded onto a HiLoad™ 26/600 Superdex™ 75 gel filtration column that had been pre-equilibrated in 100 mM sodium phosphate buffer, pH 7.4. Purified c-MYB peptides (Figure 3.2) were stored at 4°C or aliquoted, flash frozen in liquid N_2 and kept at – 80°C.

```
WT:       GSPAAAAIQRHYNDEDPEKEKRIKELELLLMSTENELKGQQVLPTQ
          NHTCSYPGW
P17A:     GSPAAAAIQRHYNDEDAEKEKRIKELELLLMSTENELKGQQVLPTQ
          NHTCSYPGW
P17/44A:  GSPAAAAIQRHYNDEDAEKEKRIKELELLLMSTENELKGQQVLATQ
          NHTCSYPGW
```

Figure 3.2 Amino Acid Sequence for Recombinant c-MYB Peptides.
c-MYB consisted of 55 amino acids: glycine and serine at the N-terminus (grey) as remnants of thrombin cleavage and residues 275 to 327 of c-MYB (red) (Uniprot P06876). Mutated residues are highlighted. The positions of mutated residues are shown relative to the peptide investigated in this work. In the context of full length c-MYB, the alanine mutations represent residues P289, and P289 and P316. The highlighted cysteine was either modified with an Alexa Fluor dye (Chapter 3.9), or mutated to an alanine to avoid disulphide bond formation.

3.6.3 PUMA

Residues 127 – 161 of mouse PUMA (UniProt Q99ML1) were expressed as a GB1 fusion protein. Residue I144 of PUMA was mutated to an alanine to reduce peptide self-association (Rogers *et al.*, 2014b) (Figure 3.3). The fusion protein gene had previously been cloned into a modified version of the pRSET A vector, which contained a HisTag N-terminal to GB1, by Dr Joseph Rogers.

```
PUMA 35: RVEEEEWAREIGAQLRRAADDLNAQYERRRQEEQH
```

Figure 3.3 Amino Acid Sequence for Recombinant PUMA 35.
The 35-amino acid PUMA peptide was expressed as a GB1-fusion protein. PUMA was cleaved from GB1 using Factor Xa, which leaves no additional residues N-terminal to the cleavage site. An I144A mutation (red) was included to reduce peptide oligomerisation. The sequence represents residues 127 – 161 of mouse PUMA (UniProt Q99ML1).

After induction with 1 mM IPTG, cells were grown in LB overnight at 18°C. Cell pellets were resuspended in PBS 25 mM imidazole and sonicated. Sonicate supernatant was incubated with Ni^{2+}-agarose resin (1.25 mL per L of culture) (Agarose Bead Technologies) for 1 hour at 4°C. Unbound protein was removed by washing twice with PBS 25 mM imidazole and once with Factor Xa cleavage buffer.

Resin was resuspended in Factor Xa cleavage buffer and incubated overnight at room temperature with 20 µg of Factor Xa (New England Biolabs): the fusion protein contained a Factor Xa cleavage site between GB1 and PUMA. Resin supernatant was loaded onto a 5 mL HiTrap™ Q HP ion exchange column (GE healthcare) using an ÄKTA FPLC (Amersham Biosciences). A TRIS pH 7 ion exchange buffer gradient of 0 – 250 mM NaCl over 90 mL was used to elute bound PUMA. Peak elution of PUMA was observed at approximately 135 mM NaCl. Fractions that appeared pure by SDS-PAGE were pooled and loaded onto a HiLoad™ 26/600 Superdex™ 30 pg gel filtration column. The column was pre-equilibrated in 50 mM biophysical sodium phosphate buffer, pH 7.0. Purified PUMA was aliquoted, flash frozen in liquid N_2 and stored at – 80°C.

3.6.4 MCL-1

Residues 152 – 308 of mouse MCL-1 (UniProt P97287) had been previously cloned into a modified version of the pRSET A vector, which contained a HisTag N-terminal to MCL-1, by Dr Joseph Rogers.

GSEDDLYRQSLEIISRYLREQATGSKDSKPLGEAGAAGRRALETLRRVGDGVQRNH
ETAFQGMLRKLDIKNEGDVKSFSRVMVHVFKDGVTNWGRIVTLISFGAFVAKHLKS
VNQESFIEPLAETITDVLVRTKRDWLVKQRGWDGFVEFFHVQDLEGG

Figure 3.4 Amino Acid Sequence for Recombinant Mouse MCL-1.
The protein sequence for MCL-1 is shown in grey. Additional glycine and serine residues (black) at the N-terminus are not part of the WT sequence, but are remnants from thrombin cleavage. The MCL-1 sequence studied in this work therefore contained these two extra residues. The N-terminal portion of MCL-1 is does not form part of the BCL-2 fold, but contains multiple post-translational modification sites (Germain and Duronio, 2007; Thomas *et al.*, 2010), while the C-terminal portion of the protein contains a trans-membrane domain. Therefore, a truncated version of MCl-1, with the first 151 and the final 23 residues removed, was investigated in this work.

After induction with 1 mM IPTG, C41 pLysS cells were grown overnight in LB at 18°C. Cell pellets were resuspended in PBS 25 mM imidazole and sonicated. Sonicate supernatant was incubated with Ni^{2+}-agarose resin (1.25 mL per L of culture) (Agarose Bead Technologies) for 1 hour at 4°C. Unbound protein was removed by washing three times with PBS 25 mM imidazole. To elute bound

protein, resin was resuspended in PBS 500 mM imidazole, which had been pH corrected to 7.5: the addition of imidazole raised the pH, taking it close to the estimated isoelectric point (PI) of MCL-1 (8.23). At the PI, a protein has no net charge, potentially increasing the likelihood that the protein will precipitate.

EDTA (4 mM final) was added to the eluted protein solution – the presence of Ni^{2+} can cause chelation of proteins with HisTags, resulting in aggregation and precipitation. EDTA binds to the nickel, preventing the HisTag chelation. To reduce the concentration of imidazole, the protein was buffer exchanged using 5 kDa molecular weight cut off (MWCO) Vivaspin centrifuge concentrators (Sartorius). Once the concentration of imidazole was reduced to ~ 50 mM, 600 units of thrombin were added to cleave the HisTag overnight at room temperature. Uncleaved protein was removed by incubation with Ni^{2+}-agarose resin. Before adding the protein to the resin, 4 mM $CaCl_2$ was added to sequester the free EDTA: free EDTA would bind to the Ni^{2+}, stripping the resin and reducing the ability of the HisTag to bind. Cleaved protein was collected from the resin supernatant. At this stage, purification of MCL-1 could be achieved by proceeding straight to size exclusion chromatography; however, some MCL-1 fractions contained an impurity of around 8 kDa (assessed by SDS-PAGE). These fractions were eluted at the start of the main MCL-1 elution peak (~ 175 mL) and appeared as a shoulder on the chromatogram. The earlier elution volume indicates that the 8 kDa protein was likely interacting with MCL-1, producing a protein species of larger size. To remove this contaminant, an ion-exchange step was included before gel filtration. Cleaved protein was loaded onto a 5 mL HiTrap™ SP HP ion exchange column (GE healthcare) using an ÄKTA FPLC (Amersham Biosciences). HEPES ion exchange buffer linear gradients of 0 – 9% over 20 mL, 9 – 13% over 25 mL and 13 – 20% over 20 mL was used to elute bound MCL-1. MCL-1 elution began at 100 mM NaCl (10%), with peak elution observed at 120 mM NaCl. Some MCL-1 also co-eluted with the 8 kDa protein at approximately 130 mM NaCl, but the majority of MCL-1 was separated from the impurity. Pure MCL-1 (assessed by SDS-PAGE) was loaded onto a HiLoad™ 26/600 Superdex™ 75 gel filtration column, which had been pre-equilibrated in 50 mM biophysical sodium phosphate buffer.

Purified MCL-1 was either stored at 4°C, or buffer exchanged into ddH_2O using HiTrap desalting columns (GE Healthcare) and lyophilised.

3.6.5 A1

The pGEX-4T-3 plasmid containing the sequence for mouse A1 was kindly donated by Bonsu Ku (Korea Research Institute of Bioscience and Biotechnology). The plasmid was a slightly modified version of pGEX-4T-3, containing a tobacco etch virus (TEV) protease site, instead of the usual thrombin cleavage site. Unlike the pRSET A plasmids used in the majority of this work, transcription of A1 was under the control of a *tac* promotor. *tac* combines the strength of the tryptophan (*trp*) promoter and the lac promotors' ability to be induced by lactose (Rosano and Ceccarelli, 2014). Therefore, protein transcription can be directly induced by the addition of IPTG, without the requirement of T7 polymerase – transcription occurs via *E.coli* RNA polymerase.

After induction with 1 mM IPTG, cells were grown in LB overnight at 24°C. Cell pellets were resuspended in PBS. A1 was expressed as a GST-fusion protein, and sonicate supernatant was incubated with glutathione sepharose 4B (GE Healthcare) for 4 hours at 4°C. Bound protein was washed once with PBS and twice with TEV cleavage buffer. After resuspending in TEV cleavage buffer, EDTA and DTT were added to a final concentration of 2 mM and 10 mM, respectively. A1 was left overnight, rolling at room temperature, with approximately 5 µM of TEV protease. Cleaved A1 was loaded onto a 5 mL HiTrap™ Q HP ion exchange column (GE healthcare) using an ÄKTA FPLC (Amersham Biosciences). A TRIS ion exchange buffer linear gradient of 0 – 23% over 60 mLs was used to elute bound A1. Elution was typically observed at approximately 100 mM NaCl (10%). After ion-exchange, A1 generally appeared pure by SDS-PAGE. As a final purification step, and to exchange the protein into biophysical buffer, A1 was loaded onto a HiLoad™ 26/600 Superdex™ 75 gel filtration column.

Purified mouse A1 consisted of residues 1-152 (Uniprot Q07440) with P104K and C113S mutations (Figure 3.5). This matched the sequence used by Smits *et al* to solve the structure of A1:PUMA and A1:BID (Smits *et al.*, 2008). The 20 residue C-terminal transmembrane domain was not included in the expressed protein. Purified A1 was stored at 4°C.

GSMAESELMHIHSLAEHYLQYVLQVPAFESAPSQACRVLQRVAFSVQKEVEKNLKS
YLDDFHVESIDTARIIFNQVMEKEFEDGIINWGRIVTIFAFGGVLLKKLKQEQIAL
DVSAYKQVSSFVAEFIMNNTGEWIRQNGGWEDGFIKKFEPKS

Figure 3.5 Amino Acid Sequence for Recombinant Mouse A1.
The protein sequence for WT A1 (UniProt Q07440) is shown in green. Additional glycine and serine residues (black) at the N-terminus are not part of the WT sequence, but are remnants from TEV cleavage. The A1 sequence studied in this work therefore contained these two extra residues.

3.6.6 BCL-XL

The pPROEX HTa plasmid containing the sequence for mouse BCL-XL was kindly donated by Bonsu Ku (Korea Research Institute of Bioscience and Biotechnology). Similar to the pGEX-4T-3 plasmid used in the expression of A1 (see Chapter 3.6.5), the pPROEX HTa plasmid does not contain a T7 promoter. Instead, protein transcription is under the control of a *trc* promoter. *trc* is based on the *tac* promoter, but contains one additional base in the linker between the *trp* and *lac* promoter segments (Brosius *et al.*, 1985). Protein transcription is therefore induced by IPTG, and facilitated by *E.coli* RNA polymerase.

BCL-XL was expressed and purified to determine its suitability as an alternative PUMA binding partner. As A1 was chosen instead of BCL-XL, the expression and purification of BCL-XL was not optimised. The protocol used to produce pure BCL-XL is described below.

After induction with 0.1 mM IPTG, cells were grown overnight at 18°C. Cell pellets were reuspended in PBS 25 mM imidazole. BCL-XL was expressed with an N--terminal HisTag. Sonicate supernatant was incubated with Ni^{2+}-agarose resin (1.25 mL per L of culture) (Agarose Bead Technologies) for 1 hour at 4°C. Unbound protein was removed by washing twice with PBS 25 mM imidazole and twice with 50 mM biophysical sodium phosphate buffer. Addition of 350 mM imidazole eluted protein from the resin. Overnight incubation at room temperature with approximately 5 µM of TEV protease and 5 mM DTT resulted in partial cleavage of the N-terminal HisTag. Size exclusion chromatography, using HiLoad™ 26/600 Superdex™ 75 gel filtration column that had been pre-equilibrated in 50 mM biophysical sodium phosphate buffer, did not separate the HisTag linked, and

cleaved BCL-XL. HisTag linked protein was removed by incubation with Ni^{2+}-agarose resin for 45 mins at room temperature. Pure BCL-XL was isolated from the resin supernatant.

Compared to the sequence of mouse BCL-XL (UniProt Q64373, residues 1 – 42 and 84 – 196), purified BCL-XL differed by a single mutation, A84S. The structure of BCL:XL (PDB 2M04) contains 5 further differences to the purified protein, an insertion (S43, which is an alanine in the mouse sequence) and 4 mutations (E40G, S84E, S168A and D193E). The disordered loop between helices 1 and 2 (residues 43 – 83) and 37 C-terminal residues were not included in the expressed protein (Figure 3.6). Purified BCL-XL was stored at 4°C.

GHMSQSNRELVVDFLSYKLSQKGYSWSQFSDVEENRTEAPEETESAVKQALREAGD
EFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDGVNWGRIVAFFSFGGALC
VESVDKEMQVLVSRIASWMATYLNDHLEPWIQENGGWDTFVDLYG

Figure 3.6 Amino Acid Sequence for Recombinant Mouse BCL-XL.
The protein sequence for BCL-XL is shown in orange. Additional glycine and histidine residues (black) at the N-terminus are not part of the WT sequence, but are remnants from TEV cleavage. The BCL-XL sequence studied in this work therefore contains these two extra residues. A disordered loop between residues 42 and 84, and the C-terminal transmembrane domain, were not included in the expressed protein.

3.7 Synthesised Peptides

3.7.1 MLL

MLL peptides were purchased from Biomatik (Canada) and synthesised with either a free, or fluorescein isothiocyanate (FITC) labelled, N–terminus (Figure 3.7). FITC is a fluorescent dye and was added to aid kinetic and thermodynamic analysis.

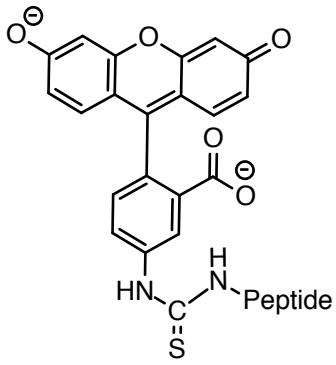


Figure 3.7 Chemical Structure of FITC.
The structure of fluorescein isothiocyanate (FITC) when attached to the N-terminus of peptides. The nitrogen-peptide bond indicates the N-terminal end of the attached peptide sequence. When conjugated to the peptide, FITC contributes an additional charge of approximately – 1.8 at pH 7 (the pKa for formation of the anion and dianion are 4.31 and 6.43, respectively) (Sjöback *et al.*, 1995). The structure of FITC was constructed using ChemDraw Professional (version 16.0).

Mouse WT MLL consisted of 31 amino acids: residues 2834 - 2865 (Uniprot P55200) with residue C2837 removed to avoid disulphide bond formation. This sequence is identical to the human version (Uniprot Q03164, residues 2838 – 2896 with C2841 removed). Single and double mutants of MLL were also purchased (P2854A; P2842A and P2854A; and L2841A) (Figure 3.8). In this work, we refer to these mutations relative to their position in the 31-residue peptide (P9A, P9/21A and L8A).

```
WT:       SDDGNILPSDIMDFVLKNTPSMQALGESPES
P21A:     SDDGNILPSDIMDFVLKNTASMQALGESPES
P9/21A:   SDDGNILASDIMDFVLKNTASMQALGESPES
L8A:      SDDGNIAPSDIMDFVLKNTPSMQALGESPES
```

Figure 3.8 Amino Acid Sequences for WT and Mutant MLL Peptides.
Peptides contained either a free, or fluorescein isothiocyanate labelled, N–terminus. The sequence for MLL WT is shown in orange. Mutated residues are highlighted. The positions of mutated residues are shown relative to the peptide investigated in this work. In the context of full length MLL, the alanine mutations represent residues P2854, P2842 and P2854, and L2841.

Peptides arrived as trifluroacetate salts and were reconstituted in biophysical 100 mM phosphate Tween buffer. Left over peptides were aliquoted, flash frozen in liquid N_2 and stored at - 80°C.

3.7.2 PUMA Shuffles

PUMA shuffles were based on the sequence of PUMA 34 (residues 128 – 161, UniProt Q99ML1). Sequences contained an M144I mutation to match the sequence used to solve the NMR structure of PUMA:MCL-1 (Day *et al.*, 2008) and the work of Dr Joe Rogers (Rogers *et al.*, 2013). The aim of this work was to create various residual structure profiles of PUMA 34 by moving the relative position of the residues. As only the position of the residue is altered, the overall sequence composition remains the same. Effects on the kinetics and thermodynamics of the shuffled peptides binding to MCL-1 would then be monitored. Clearly, altering the position of contacting residues could lead to a significant destabilisation of the bound complex. Therefore, 7 hydrophobic interface residues were kept constant (W133, I137, L141, I144, A145, L148 and Y152) (Figure 3.9). The choice of these 7 residues was a balance between maintaining the key contacts, while leaving enough residues to shuffle.

```
WT:     VEEEEWAREIGAQLRRIADDLNAQYERRRQEEQH
BH1:    EVDEEWQARIQRALQGIAEDLNARYRERRQEEEH
BH2:    EVDEEWARGIRAQLQRIAEDLNARYRERQEEQEH
MSH1:   EGREEWEEDIAQVLARIAHQLRENYERREDAQRQ
MSH2:   ERDEEWREDIEEALAVIARQLNQQYGQRERRAEH
NTH1:   EVEEEWARGIRARLQRIAEDLNARYERQQDEQEH
NTH2:   GEVDEWAQRIRAQLHRIAEDLNEAYRREREQEQE
CTH1:   EVEEQWQAQIEAGLREIADDLNARYRRRHRQEEE
CTH2:   EVDERWEARIQEALQGIAEDLNAEYRRRRQEQEH
LSH1:   VREEEWERAIRGQLREIAQDLNQDYRERAAEQEH
LSH2:   RVEEGWAQRIEQRLEAIADDLNREYAERRQQEEH
```

Figure 3.9 Amino Acid Sequences for PUMA 34 Shuffles.
Residues of WT PUMA 34 were shuffled in an attempt to create various residual structure profiles (BH – broken helix; MSH – middle spanning helix; NTH – N-terminal helix; CTH – C-terminal helix; LSH – low stability helix). Two versions of each design were purchased. The position of 7 PUMA:MCL-1 interface residues were kept constant in all designs (blue). Peptides were N-terminally acetylated and C-terminally amidated.

This work was performed in collaboration with Prof. Rohit Pappu's group (University of Washington in St. Louis). Tyler Harmon, a PhD student in Prof. Pappu's lab, generated the shuffle sequences using their algorithm, GADIS (Harmon *et al.*, 2016). In essence, the Genetic Algorithm for Design of Intrinsic secondary Structure (GADIS) first creates shuffled sequences. Input parameters are the peptide sequence, the position of immutable residues and the target helicity profile. The shuffled sequences are then simulated, yielding conformational ensembles. Comparison with the target helicity profile provides an objective score. Based on this score, sequences are chosen, reshuffled, simulated and scored. Sequences are selected once they go through two generations without improving their score. Figure 3.9 represents the shuffled sequences of PUMA 34 generated by GADIS. For reasons described in Chapter 5.3, 5 additional peptides were designed. Briefly, these consisted of 3 shuffles with 19 PUMA residues fixed, and 2 shuffles based on single site swaps of WT and CTH1 (Figure 3.10).

CTH 1:	EVEEQWQAQIEAGLREIADDLNARYRRRHRQEEE
CTH G Swap:	EVEEQWQAQI**G**A**E**LREIADDLNARYRRRHRQEEE
WT G Swap:	VE**G**EEWAREI**E**AQLRRIADDLNAQYERRRQEEQH
WT:	VEEEEWAREIGAQLRRIADDLNAQYERRRQEEQH
Conservative 1:	EQEEEWAREIGAQLRRIADDLNAQYERRRQEEQH
Conservative 2:	EVQEQWAREIGEQLREIADELNRAYEQRRREDAH
Conservative 3:	EREEEWAREIGRQLRDIADHLNQVYEARRAEEQQ

Figure 3.10 Amino Acid Sequences for Additional PUMA 34 Shuffles.
On a background of CTH 1, a glycine was swapped with a glutamic acid (red), returning the glycine to the position occupied in WT PUMA (CTH G Swap). A glutamic acid (red) was swapped into the same position in the context of the WT PUMA 34 sequence, at the expense of glycine (WT G Swap). Keeping 19 residues in their WT positions (blue), 3 conservative swaps were designed. A contact map analysis (Sobolev *et al.*, 2005) of the PUMA:MCL-1 NMR structure (PDB:2ROC) was used to determine the residues kept constant. The 19 residues identified all had a contact area of 11 Å or greater.

N-terminal acetylated and C-terminal amidated peptides were purchased from WatsonBio (Texas, US). Peptides arrived as trifluroacetate salts and were reconstituted in biophysical 50 mM phosphate Tween buffer. To ensure the removal

of residual salts, peptides were buffer exchanged into biophysical 50 mM phosphate Tween buffer using 5 mL Hitrap™ desalting columns (GE Healthcare).

3.7.3 BIM, BIK and PUMA Core Swaps

To assess the impact of non-contacting residues on the kinetics and thermodynamics of PUMA binding to MCL-1, variants of mouse PUMA containing either WT PUMA, BIM or BIK amino acids were designed. As for the PUMA 34 conservative shuffles (Figure 3.10), the 19 residues of PUMA that contact MCL-1 were kept constant. The remaining 16 PUMA residues were swapped with residues from BIM (UniProt O54918) or BIK (UniProt O70337).

BIK:	LMEAVEGRNQVALRLAAIGDEMDLALRSPRLVQLP
BIK Swap:	**LMEA**E**E**WAREIG**L**QLR**A**AAD**E**LN**LA**YE**S**RR**LVQLP**
PUMA:	RVEEEEWAREIGAQLRRAADDLNAQYERRRQEEQH
BIM Swap:	**EPED**E**R**WAREIG**Q**QLR**R**AAD**E**LN**ET**YE**R**RR**FANDY**
BIM:	EPEDLRPEIRIAQELRRIGDEFNETYTRRVFANDY

Figure 3.11 Amino Acid Sequences for PUMA Core Swaps.
Oligomerisation problems had been observed with the PUMA 34 shuffles (Chapter 5.3.1). To reduce the propensity of the peptide to self-associate, an I144A mutation (Rogers *et al*., 2014b) was included in the design of these peptides. Differences in net charge between PUMA and the swap peptides could alter long-range electrostatic interactions with MCL-1, causing a change in the association rate constant. Therefore, PUMA 35, which contains an additional arginine at the N-terminus, was used as the template sequence. This allowed the PUMA, BIK-swap and BIM-swap to maintain estimated net charges of approximately – 3. Based on the number of positively and negatively charged residues, BIM and BIK have estimated net charges of – 3 and – 1, respectively. To avoid disulphide bond formation, 3 cysteine residues in BIK were mutated to alanine (C44A, C57A and C65A). PUMA residues are shown in blue, swapped residues are indicated in black.

N-terminal acetylated and C-terminal amidated peptides were purchased from WatsonBio (Texas, US). Peptides arrived as trifluroacetate salts and were reconstituted in biophysical 50 mM phosphate Tween buffer. To ensure the removal of residual salts, peptides were buffer exchanged into biophysical 50 mM phosphate Tween buffer using 5 mL HiTrap™ desalting columns (GE Healthcare).

3.7.4 PUMA TAMRA Peptides

Variants of residues 127 – 161 of mouse PUMA (UniProt Q99ML1) were purchased from Biomatik (Canada) (Figure 3.12).

WT:	RVEEEEWAREIGAQLRRAADDLNAQYERRRQEEQH
E132A:	RVEEE**A**WAREIGAQLRRAADDLNAQYERRRQEEQH
E132G:	RVEEE**G**WAREIGAQLRRAADDLNAQYERRRQEEQH
W133F:	RVEEEE**F**AREIGAQLRRAADDLNAQYERRRQEEQH
E136A:	RVEEEEWAR**A**IGAQLRRAADDLNAQYERRRQEEQH
E136G:	RVEEEEWAR**G**IGAQLRRAADDLNAQYERRRQEEQH
I137A:	RVEEEEWARE**A**GAQLRRAADDLNAQYERRRQEEQH
A139G:	RVEEEEWAREIG**G**QLRRAADDLNAQYERRRQEEQH
L141A:	RVEEEEWAREIGAQ**A**RRAADDLNAQYERRRQEEQH
R143A:	RVEEEEWAREIGAQLR**A**AADDLNAQYERRRQEEQH
R143G:	RVEEEEWAREIGAQLR**G**AADDLNAQYERRRQEEQH
D147A:	RVEEEEWAREIGAQLRRAAD**A**LNAQYERRRQEEQH
D147G:	RVEEEEWAREIGAQLRRAAD**G**LNAQYERRRQEEQH
L148A:	RVEEEEWAREIGAQLRRAADD**A**NAQYERRRQEEQH
A150G:	RVEEEEWAREIGAQLRRAADDLN**G**QYERRRQEEQH
Y152A:	RVEEEEWAREIGAQLRRAADDLNAQ**A**ERRRQEEQH
R154A:	RVEEEEWAREIGAQLRRAADDLNAQYE**A**RRQEEQH
R154G:	RVEEEEWAREIGAQLRRAADDLNAQYE**G**RRQEEQH
A139G A150G:	RVEEEEWAREIG**G**QLRRAADDLN**G**QYERRRQEEQH

Figure 3.12 Amino Acid Sequences for Synthesised PUMA TAMRA Peptides. The sequence for WT PUMA 35 is shown in blue. Compared to PUMA 34, PUMA 35 contains an additional arginine at the N-terminus and an I144A mutation. These two changes were shown to reduce the self-association propensity of PUMA 34. (Rogers *et al*., 2014b) Investigated mutations are highlighted in black. Peptides were synthesised with fluorescent dye, TAMRA, attached to the N-terminus via a peptide bond.

As discussed in Chapter 8, a fluorescent dye was required to follow the binding of PUMA to the BCL-2 like protein, A1. Dissociation of PUMA from MCL-1, another BCL-2 like protein, had been shown previously to be relatively slow (k_{off} = 0.0013 ± 0.0009 s^{-1}) (Rogers *et al*., 2014a). With this rate, measuring 10 half-lives would require monitoring the reaction for approximately 5,000 s. Over such timescales, fluorophores may photo-bleach. A balance between price and photo-stability

(Dempsey *et al.*, 2011) therefore led to 5-Carboxytetramethylrhodamine (TAMRA) (Figure 3.13) being chosen as a good candidate for this work.

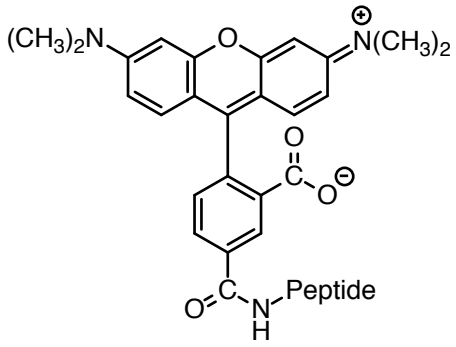


Figure 3.13 Chemical Structure of TAMRA.
The structure of 5-Carboxytetramethylrhodamine (TAMRA) when attached to the N-terminus of peptides. The nitrogen-peptide bond indicates the N-terminal end of the attached peptide sequence. Conjugation of TAMRA does not change the net charge of the peptide. ChemDraw Professional (version 16.0) was used to construct the structure of TAMRA.

The 35 amino acid peptides were N-terminally labelled with TAMRA. Peptides arrived as trifluroacetate salts and were reconstituted in biophysical 50 mM phosphate Tween buffer. Left over peptides were aliquoted, flash frozen in liquid N_2 and stored at - 80°C.

3.7.5 BID

Residues 76 – 110 of mouse BID (UniProt P70444) were purchased from Biomatik (Canada) (Figure 3.14).

WT BID: SESQEEIIHNIARHLAQIGDEMDHNIQPTLVRQLA

Figure 3.14 Amino Acid Sequence of WT BID.
The peptide was synthesised with a fluorescent dye, TAMRA, attached to the N-terminus via a peptide bond.

To allow comparison with PUMA TAMRA, the 35-amino acid peptide was N-terminally labelled with the fluorescent dye, TAMRA (Figure 3.13). The peptide

arrived as a trifluroacetate salts and was reconstituted in biophysical 50 mM phosphate Tween buffer. Left over peptide was aliquoted, flash frozen in liquid N_2 and stored at - 80°C.

3.8 Isotopically Labelled Proteins

3.8.1 c-MYB

Expression and purification of ^{13}C ^{15}N isotopically labelled c-MYB was performed by Dr Wade Borcherds, a post-doc in Dr Gary Daughdrill's lab (University of South Florida). Except for a couple of nuances, it was carried in the same manner as non-isotopically-labelled c-MYB (Chapter 3.6.2). Firstly, to avoid disulphide bond formation, the cysteine at position 50 (C322 in the full-length protein) was mutated to an alanine. Secondly, to produce isotopically labelled protein, C41 DE3 *Escherichia coli* cells were grown to an OD_{600} of 0.4 – 0.6 at 37°C in minimal media containing M9 salts and 100 mg/L ampicillin, with ^{13}C D-glucose and ^{15}N ammonium chloride (Cambridge Isotope Labortories) as the sole carbon and nitrogen sources.

3.8.2 MLL

Residues 2834 - 2865 (Uniprot P55200) with residue C2837 removed to avoid disulphide bond formation were expressed as a GB1 fusion protein. The fusion protein gene had previously been cloned into a modified version of the pRSET A vector, which contained a HisTag N-terminal to GB1, by Dr Sarah Shammas.

C41 DE3 Escherichia coli cells were grown to an OD600 of 0.4-0.6 at 37°C in minimal media containing M9 salts and 100 mg/L ampicillin, with 13C D-glucose and 15N ammonium chloride (Cambridge Isotope Labortories) as the sole carbon and nitrogen sources. Protein expression was induced by addition of 1 mM isopropyl-beta-D-thiogalactopyranoside (IPTG). After incubating for 4 hours at 37°C, cells were harvested by centrifugation. Cells were re-suspended in PBS, 25 mM imidazole, sonicated and centrifuged. His-GB1-MLL was purified from the soluble fraction by affinity chromatography using a 5 mL HisTrap column (GE Healthcare Life Sciences). Bound protein was eluted using a linear gradient of PBS 0 - 500 mM imidazole over 50 mL: elution began at approximately 50 mM imidiazole. To reduce the ionic strength, protein was diluted 5-fold before loading onto a HiTrap Q HP ion exchange column (GE Healthcare Life Sciences). His-GB1-

MLL was separated from truncated versions using a linear gradient of 20 mM TRIS pH 7.0, 1 M NaCl (0 – 300 mM NaCl over 50 mL). Full length His-GB1- MLL eluted at ~ 200 mM NaCl. His-GB1-MLL was bound to Ni^{2+} agarose resin and exchanged into 100 mM sodium phosphate pH 7.4. MLL was cleaved from His-GB1 by overnight incubation with 100 units of thrombin. Thrombin was removed from the sample by filtration using a 20 kDa cut-off filter (GE Healthcare Life Sciences), with the flow through containing MLL. The identity and purity of MLL were confirmed by mass spectrometry (Department of Chemistry, University of Cambridge, UK) and amino acid analysis (Department of Biochemistry, University of Cambridge, UK).

3.9 Protein and Peptide Dye Labelling

To reduce disulphide bonds between cysteine residues, purified peptides were incubated with 1 mM tris(2-carboxyethyl)phosphine) (TCEP) for at least 30 minutes at 37°C. Alexa Fluor® 594 C5 maleimide dye (Life Technologies) was added to the peptides at molar a ratio of 1.4:1 and the solution left covered, rolling at room temperature for 2 hours. The solution was diluted with ddH_2O to reduce the ionic strength and β-mercaptoethanol was added to a final concentration of 10 mM to quench the reaction. Labelled peptide was separated from unlabelled peptide and free dye by ion exchange: the solution was loaded onto a HiTrap Q ion exchange column (GE healthcare) and separated using TRIS pH 7 ion exchange buffer with a gradient of 0 to 1M NaCl over 90 mL. Purified labelled peptides were exchanged in biophysical phosphate buffer using a HiTrap desalting column (GE Healthcare). Labelled peptides were stored at 4°C or aliquoted, flash frozen in liquid N_2 and kept at –80°C.

3.10 Protein and Peptide Concentration Determination

Knowledge of accurate protein concentrations is crucial when conducting a biophysical analysis of bimolecular interactions, due to the concentration dependent nature of the reaction. Substantial care was therefore taken throughout this work to ensure that protein concentrations were determined as accurately as possible.

Protein and peptide stock concentrations were typically determined using a Cary 60 UV-Vis spectrophotometer (Agilent Technologies) and the Beer – Lambert Law, as shown in Equation 3.1:

$$A = \varepsilon c l \tag{3.1}$$

where A is the absorbance at a specific wavelength, ε is the extinction coefficient at that wavelength, c is the concentration and l is the path length.

Before calculating the concentration, a buffer absorbance reading was subtracted from all protein and peptide absorbance scans. When necessary, protein and peptide stocks were diluted with buffer to achieve an absorbance reading of between 0.1 and 1 absorbance units.

3.10.1 Extinction Coefficients

Protein absorbance at 280 nm is predominantly due to the presence of tryptophan, tyrosine and cystine residues. Estimates of a protein's extinction coefficient can thus be calculated by multiplying the number of tryptophan, tyrosine and cystine residues by their molar extinction coefficients (5500 M^{-1} cm^{-1} for tryptophan, 1490 M^{-1} cm^{-1} for tyrosine and 125 M^{-1} cm^{-1} for cystine) (Pace *et al.*, 1995). However, this approach does not take into account the positioning or context of these residues and, therefore, does not necessarily yield an extinction coefficient that is accurate enough to determine the true protein concentration. For example, the predicted extinction coefficient for the folded BCL-2 protein, A1, is 22460 M^{-1} cm^{-1}, compared to the experimentally determined value of 24200 M^{-1} cm^{-1}. Ergo, use of the predicted extinction coefficient would lead to an overestimation of the concentration of A1 by more than 7%.

Proteins can also be associated with other chromophores. For example, to aid in biophysical analysis, many of the proteins and peptides in this work were labelled with fluorescent dyes. As they are designed to absorb and fluoresce light, these chromophores typically have large extinction coefficients, which can be utilised to determine protein concentrations. Many of the dyes also absorb partially at 280 nm, precluding the use of extinction coefficients that are calculated purely from the

amino acid content. As described in the subsections below, several methods were consequently employed to confirm extinction coefficients and protein concentrations.

3.10.1.1 CBP KIX

An extinction coefficient of 12,009 M^{-1} cm^{-1}, which had been previously determined by Dr Sarah Shammas (Shammas *et al.*, 2013) using the method of Gill von Hippel (Gill and von Hippel, 1989), was used for CBP KIX.

3.10.1.2 MLL and c-MYB

The extinction coefficient of 8,480 M^{-1} cm^{-1} at 280 nm for unlabelled c-Myb was estimated using the number of tryptophan, tyrosine and cystine residues and their respective determined extinction coefficients (Pace *et al.*, 1995). For peptides labelled with a dye, the extinction coefficient of the dye was used: extinction coefficients for Alexa Fluor® 594 labelled c-Myb and FITC labelled MLL were 92,000 M^{-1} cm^{-1} at 593 nm and 75,000 M^{-1} cm^{-1} at 493 nm, respectively. Determined protein concentrations were checked by amino acid analysis (Department of Biochemistry, University of Cambridge), which provides an estimate that is suggested to be within 5% of the actual concentration.

3.10.1.3 MCL-1 and PUMA 35

Dr Joseph Rogers had previously determined the extinction coefficients for MCL-1 (Rogers *et al.*, 2013) and PUMA 35 (Rogers *et al*., 2014a) through comparison of the absorbance at 280 nm with the concentration determined from amino acid analysis. These extinction coefficients of 22157 M^{-1} cm^{-1} and 7113 M^{-1} cm^{-1} for MCL-1 and PUMA 35, respectively, were used in this work.

3.10.1.4 PUMA Shuffles and Core Swaps

Samples of each stock solution were weighed to ascertain an accurate volume. Concentrations were then determined by amino acid analysis (Department of Biochemistry, University of Cambridge).

3.10.1.5 PUMA-TAMRA and BID-TAMRA Peptides

Solutions of WT and mutant PUMA and BID TAMRA peptides were scanned using a Cary 60 UV-Vis spectrophotometer (Agilent Technologies). An absorbance peak at 555 nm was observed and buffer subtracted. A range of stock dilutions were generated to produce values from 0.1 to 1 absorbance units. Samples of the scanned stocks were sent for amino acid analysis (Department of Biochemistry, University of Cambridge). Extinction coefficients were calculated from the gradient of the straight line fit of the absorbance at 555 nm verses the concentration from amino acid analysis (Figure 3.15). For PUMA-TAMRA, several of the mutants displayed a propensity to oligomerise at concentrations within the micro molar range (see Chapter 8.1.3). Oligomerisation was found to influence the absorbance at 555 nm and, consequently, oligomerising peptides were not included in the determination of the extinction coefficient. Values of 83,000 ± 1,000 M^{-1} cm^{-1} and 85,000 ± 1,000 M^{-1} cm^{-1} at 555 nm were established and utilised for PUMA-TAMRA and BID-TAMRA, respectively.

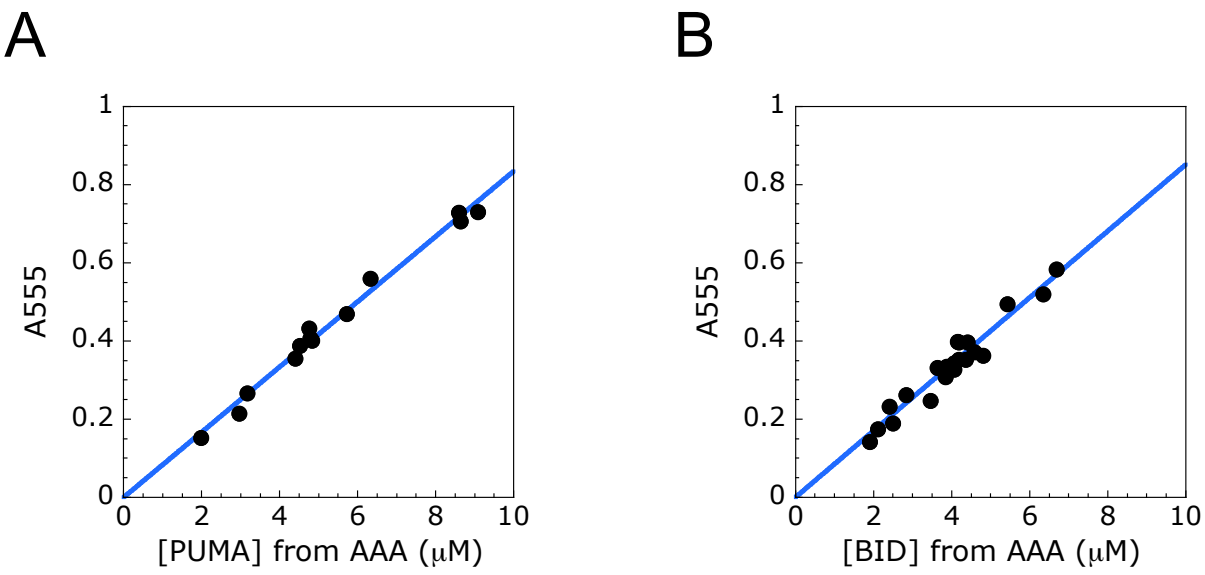


Figure 3.15 Extinction Coefficient for TAMRA labelled Peptides.
Samples of PUMA-TAMRA and BID-TAMRA peptides were scanned and the absorbance at 555 nm was recorded and buffer subtracted (A555). Concentrations for the samples were determined by amino acid analysis (AAA). Gradients of the straight-line fits gave extinction coefficients of 83,000 ± 1,000 M^{-1} cm^{-1} for PUMA-TAMRA (A) and 85,000 ± 1,000 M^{-1} cm^{-1} for BID-TAMRA (B). For fitting, the intercept was fixed to 0 — when no protein is present, the absorbance should be equal to the absorbance of the buffer. The reported errors represent the error of the fit.

3.10.1.6 A1

A1 was expressed and purified on several occasions during this work. To experimentally determine an extinction coefficient for A1, a buffer subtracted absorbance at 280 nm was obtained. Concentrations for a number of the scanned solutions were ascertained from amino acid analysis (Department of Biochemistry, University of Cambridge). A straight-line fit of the absorbance at 280 versus the concentration provided an extinction coefficient of 24,200 ± 200 M^{-1} cm^{-1} (Figure 3.16). This extinction coefficient was then utilised for the remaining preparations that had not been sent for amino acid analysis.

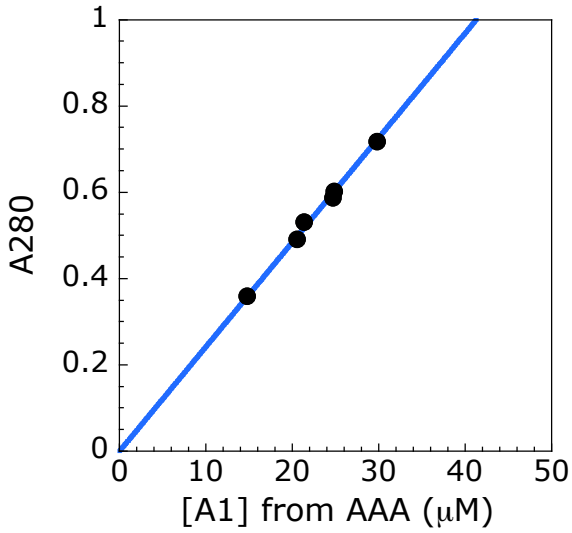


Figure 3.16 A1 Extinction Coefficient
Samples of A1 were scanned and the absorbance at 280 nm was recorded and buffer subtracted (A280). Concentrations for the various A1 preparations were determined by amino acid analysis (AAA). An extinction coefficient of 24,200 ± 200 M^{-1} cm^{-1} was obtained from the gradient of the straight-line fit. When no protein is present, the absorbance should be equal to the absorbance of the buffer; therefore, the intercept of the fit was fixed to a value of 0. The reported error represents the error of the fit.

3.10.1.7 BCL-XL

The folded BCL-2 like protein, BCL-XL, was expressed and purified to test its suitability as an alternative to MCL-1. After initial experiments, A1 was chosen for further study, resulting in little biophysical analysis being pursued with BCL-XL. An experimentally determined extinction coefficient was, therefore, not acquired.

Instead, extinction coefficients were simply calculated by multiplying the number of tryptophan, tyrosine and cystine residues by their molar extinction coefficients.

3.11 Circular Dichroism

Circular dichroism (CD) experiments were performed using Applied Photophysics Chirascan spectrometers. To ensure accurate concentrations, dilutions of stock solutions were produced by weight. Spectra obtained with buffer were subtracted from the protein spectra.

The millidegrees (mdeg) signal obtained from far-UV CD is related to the number of peptide bonds, and the secondary structure of the protein. Thus, the raw millidegrees signal from one protein of a given length and concentration cannot be directly compared to the signal obtained with another protein of different length and/or concentration. To allow comparison, the CD signal must be converted to mean residual ellipticity (MRE), which takes into account the protein concentration, cuvette path length and the number of amino acids (aa). This conversion was achieved using Equation 3.2 and gives the MRE with units of degrees cm^2 $dmol^{-1}$.

$$\mathrm{MRE} = \frac{\mathrm{CD\ signal\ (mdeg)}}{(10 \times \mathrm{concentration\ (M)} \times \mathrm{path\ length\ (cm)} \times \mathrm{no.of\ aa})} \qquad (3.2)$$

The path length is included in Equation 3.2 due to the relationship between the number of peptide bonds and the observed CD signal: at the same concentration, a longer path length cuvette would result in the polarised light interacting with more protein molecules. Thus, giving a greater mdeg signal. However, increasing the path length leads to a higher background absorbance. The reduced amount of transmitted light can be offset by enhancing the detector voltage. Although, this comes at the cost of greater signal noise. For the Applied Photophysics instruments utilised in this work, a detector voltage of 800 V was the recommended maximum. A delicate balance of protein concentration and path length is therefore required to optimise the signal to noise ratio. Details for the CD settings employed are listed in the subsections below.

3.11.1 MLL, c-MYB and CBP KIX

Association of CBP KIX to its intrinsically disordered partners was known to be rapid (Shammas *et al.*, 2013). Experiments with CBP KIX were therefore performed at low temperatures (10°C for c-Myb:KIX and 5°C for MLL:KIX). Cuvettes of 1 or 2 mm path lengths were utilised. Settings were 1 nm bandwidth and 5 s adaptive sampling. The mean average of three technical repeats was used to calculate helicity estimates.

3.11.2 MCL-1, PUMA Shuffles and Core Swaps

PUMA 34, the template for the PUMA shuffles, self-associates in the micro-molar range. Upon oligomerising, PUMA 34 gains helical structure, producing an estimated helicity that is concentration dependent (Rogers *et al.*, 2013). To ensure that the calculated helicities were concentration independent and representative of the monomeric peptide, at least 3 concentrations (typically 2.5, 5 and 10 μM) were compared. MCL-1 samples were scanned at a concentration of 5 μM. A 2 mm cuvette was utilised to ensure that recorded signals achieved a reasonable signal to background ratio. Settings were 1 nm bandwidth, 25°C and 15 s adaptive sampling. For each concentration of peptide and MCL-1, two samples were independently prepared and scanned. Estimated helicities were calculated from the mean of the highest concentration samples that showed no concentration dependence.

3.11.3 A1

A1 contains a single cysteine. To prevent disulphide-bond formation, a reducing agent (10 mM DTT) was added. Somewhat problematically, DTT absorbs in the far-UV region, increasing the background absorbance. A 0.2 mm cuvette was therefore used to allow enough light to reach the detector. Settings were 1 nm bandwidth, 25°C and 5 s adaptive sampling. A ten-fold concentration range (4, 10, 20, 30 and 40 μM) was compared to check for self-association.

3.11.4 PUMA TAMRA Peptides

Addition of an arginine at the N-terminus in PUMA 35 and an I144A mutation had reduced the self-association problems observed with PUMA 34.(Rogers *et al*., 2014b; Rogers *et al*., 2013) However, the addition of TAMRA, a hydrophobic dye,

resulted in the formation of PUMA 35 oligomers. Low concentrations of PUMA-TAMRA peptides (typically 0.5 and 1 μM) were consequently scanned using a cuvette of 1 cm path length. Use of this cuvette resulted in an increased background absorbance. To prevent the detector voltage exceeding 800 V, scans were stopped at 200 nm. Settings were 2 nm bandwidth, 25°C and 15 s adaptive sampling. All peptides were reconstituted and scanned on the same day. Peptide concentrations were determined from amino acid analysis (Department of Biochemistry, University of Cambridge). Helicity estimates were calculated from the highest concentration sample that showed no concentration dependence.

3.11.5 Coupled Folding and Binding

Upon binding, a subset of IDPs are known to fold into a well-defined structure. This folding and binding reaction can be probed using CD. Spectra for near equimolar concentrations of each individual sample were collected. The mean of the two individual sample spectra was calculated and converted to MRE: this represents the MRE spectrum expected if no change in structure occurred upon binding. A mixture of the two samples was produced in a 1:1 volume ratio and scanned. The difference between the observed spectrum and the mean of the individual spectra is related to the amount of structural change and the proportion of protein in complex. Therefore, to allow comparison of the structure formation upon binding, the difference spectra were multiplied by the fraction of protein in complex.

The concentration of complex was estimated using Equation 3.3:

$$[AB] = \frac{\left([A] + [B] + K_d - \sqrt{(([A] + [B] + K_d) - 4[A][B])}\right)}{2} \qquad (3.3)$$

where $[A]$ is the total concentration of protein A, $[B]$ is the total concentration of protein B and K_d is the dissociation constant for the reaction.

The fraction bound was then calculated by dividing the concentration of complex ($[AB]$) by the total amount of either protein A or protein B.

3.11.6 Estimation of Helical Content

Helical content was estimated from the MRE value at 222 nm using the method of Muñoz and Serrano (Muñoz and Serrano, 1995). Their method is described below.

The expected MRE for a fully helical protein at 0°C is given by Equation 3.4:

$$\mathrm{Helix}_{0^{\circ}\mathrm{C}} = -39{,}500\left(1 - \frac{2.57}{n}\right) \tag{3.4}$$

where $\mathrm{Helix}_{0^{\circ}\mathrm{C}}$ is the expected MRE for a full helix at 0°C and n is the number of amino acids.

The MRE values for a random coil at a certain temperature is determined using Equation 3.5:

$$\mathrm{MRE}_{\mathrm{Coil}} = 400 - 45t \tag{3.5}$$

where 400 is the MRE at 222 nm for a random coil at 0°C and t is the temperature in degrees Celsius.
Equation 3.6 gives the MRE for a full α-helix at a specific temperature.:

$$\mathrm{MRE}_{\mathrm{Helix}} = \mathrm{Helix}_{0^{\circ}C} + 100t \tag{3.6}$$

where $\mathrm{Helix}_{0^{\circ}\mathrm{C}}$ is given by Equation 3.4 and t is the temperature in degrees Celsius.

Muñoz and Serrano's method then estimates the helicity using Equation 3.7:

$$\%\text{Helix} = \frac{100}{\left(1+\left(\frac{(\text{MRE}_{222}-\text{MRE}_{\text{Helix}})}{(\text{MRE}_{\text{Coil}}-\text{MRE}_{222})}\right)\right)} \quad (3.7)$$

where MRE_{222} is the measured mean residual ellipticity at 222 nm, MRE_{Coil} is given by Equation 3.5 and MRE_{Helix} is given by Equation 3.6.

3.12 Nuclear Magnetic Resonance

The nuclear magnetic resonance (NMR) data presented in Chapter 4 was collected in collaboration with Dr Gary Daughdrill's Lab (University of South Florida, US). Isotopically labelled MLL protein was expressed, purified and shipped to Dr Daughdrill. Expression and purification of ^{13}C ^{15}N c-MYB was performed in Florida by Dr Wade Borcherds, a post-doc in Dr Daughdrill's lab. Dr Borcherds and Dr Daughdrill performed the NMR experiments. Dr Borcherds, Dr Daughdrill and I analysed the data.

NMR experiments were performed on a Varian VNMRS 800 MHz spectrometer equipped with a triple resonance pulse field *Z*-axis gradient cold probe. To make the amide ^{1}H and ^{15}N as well as $^{13}C_{\alpha}$, $^{13}C_{\beta}$ and ^{13}CO resonance assignments, sensitivity enhanced ^{1}H-^{15}N HSQC and three-dimensional HNCACB and HNCO experiments were performed on uniformly ^{15}N and ^{13}C-labelled samples. All NMR spectra were processed with NVFx and analysed using nmrViewJ software (Johnson, 2004; Johnson and Blevins, 1994). The ^{1}H carrier frequency was set on 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) and used as the reference frequency (Markley *et al.*, 1998). Secondary chemical shift values were calculated by subtracting the residue specific random coil chemical shifts in the neighbour-corrected IDP chemical shift library (ncIDP), from the measured chemical shifts (Tamiola *et al.*, 2010). Secondary structure populations were calculated with the δ2D method, using the measured proton, nitrogen, and alpha, beta, and carbonyl carbon chemical shifts (Camilloni *et al.*, 2012). Overall helicity was calculated as the mean of the per residue δ2D helical population estimates.

3.12.1 c-MYB

NMR experiments for c-MYB and its proline to alanine mutants were performed with 200 μM of uniformly ^{15}N- and ^{13}C-labelled samples at both 10°C and 25°C. Experiments were carried out in 90% H_2O/10% D_2O, 50mM sodium phosphate buffer with 50 mM NaCl, 1mM EDTA, 0.02% NaN_3, at pH 6.8. For WT c-MYB, processing and analysis of the HNCACB data resulted in 48 non-proline, amide ^{1}H, ^{15}N, $^{13}C_\alpha$ and 48 $^{13}C_\beta$ resonance assignments plus 4 proline $^{13}C_\alpha$ and $^{13}C_\beta$ resonance assignments. 47 ^{13}CO resonance assignments were made using the HNCO data. For P17/44A c-MYB, 46 non-proline, amide ^{1}H, ^{15}N, $^{13}C_\alpha$ and 43 $^{13}C_\beta$ resonance assignments plus 2 $^{13}C_\alpha$ and 2 $^{13}C_\beta$ proline resonance assignments were made using the HNCACB data and 46 ^{13}CO resonance assignments were made using the HNCO data.

3.12.2 MLL

NMR experiments for WT MLL and MLL P9/21A were performed at 10°C using sample concentrations of 160 μM, and 155 μM, respectively. Experiments were carried out in 90% H_2O/10% D_2O, 50mM sodium phosphate buffer with 50mM NaCl, 1mM EDTA, 0.02% NaN_3, at pH 6.8. For WT MLL, processing and analysis of the HNCACB data resulted in 28 non-proline, amide ^{1}H, ^{15}N, $^{13}C_\alpha$ and 26 $^{13}C_\beta$ resonance assignments plus 3 proline $^{13}C_\alpha$ and $^{13}C_\beta$ resonance assignments. Using the HNCO data, 27 ^{13}CO resonance assignments were made. For mutant MLL, processing and analysis of the HNCACB data resulted in 30 non-proline, amide ^{1}H, ^{15}N, $^{13}C_\alpha$ and 28 $^{13}C_\beta$ resonance assignments plus 1 proline $^{13}C_\alpha$ and $^{13}C_\beta$ resonance assignments. 29 ^{13}CO resonance assignments were made using the HNCO data.

3.13 Association Kinetics

Protein solutions were rapidly mixed using Applied Photophysics SX18 or SX20 stopped-flow spectrometers. Temperature was maintained at 10°C for c-MYB, 5°C for MLL and 25°C for BCL-2 family experiments. Depending on the fluorophore, excitation wavelengths of 593 nm (a594-c-MYB, Chapter 4), 493 nm (FITC-MLL, Chapter 4), 280 nm (PUMA:MCL, Chapters 5 and 6) and 555 nm (TAMRA-Peptides, Chapters 7 and 8) were utilised. To reduce the detection of scattered light, long-pass filters were employed between the cell and photomultiplier tube: 515 nm for FITC-MLL, 610 nm for a594-c-MYB, 320 nm for PUMA:MCL-1 and 570 nm for TAMRA-labelled peptides. Depending on the photobleaching propensity of the

proteins and the timescale of the reaction, slit widths were set to either 1 or 2 mm. Except for MLL P9/21A and L8A (See Chapter 3.15), fluorescence traces were collected and averaged before analysis. Data within the deadtime of mixing (the first 1 ms) were removed before fitting.

3.13.1 Pseudo-First-Order Association Kinetics

Solutions were prepared such that the concentration of one protein was at least 10-fold higher than the concentration of the partner. To ensure optimal signal to noise, the protein that had the lower relative fluorescence was used in excess (KIX for Chapter 4 and the BCL-2 folded partner for Chapters 7 and 8). For each concentration of excess protein, the mean trace was fit to a single exponential decay function (Equation 2.25) to extract the observed rate of reaching the new equilibrium (k_{obs}). k_{obs} is equal to Equation 3.8:

$$k_{\mathrm{obs}} = k_{\mathrm{on}}[A] + k_{\mathrm{off}} \tag{3.8}$$

where [A] is the concentration of protein in excess, k_{on} is the association rate constant and k_{off} is the dissociation rate constant.

Therefore, the gradient from the straight line fit of k_{obs} vs the concentration of excess protein was used to determine k_{on}.

3.13.2 Association Kinetics for PUMA Shuffles and Core Swaps

Association kinetics were monitored by following the change in intrinsic tryptophan fluorescence upon binding. MCL-1 (0.2 μM) was rapidly mixed with various concentrations of PUMA (0.2 – 1 μM). The voltage applied to the photomultiplier tube was kept constant for an individual peptide and the data for all concentrations were collected on the same day. Traces for each concentration were averaged (mean), before fitting to a reversible biomolecular association model (Equation 2.26), to determine the k_{on}. When the K_d fell within a suitable range (>~1 nM), the K_d could also be reliably determined from this model. All concentrations were globally fit using ProFit (Quantum Soft), with the k_{on}, K_d and fluorescence change shared. The starting fluorescence was fit individually for each concentration, while

the concentration of MCL-1 and the PUMA to MCL-1 ratio were fixed. For peptides that oligomerised, a linear term was added to the model.

3.14 Dissociation Kinetics

Depending upon the observed reaction rate, dissociation kinetics were monitored using either Applied Photophysics SX18 or SX20 stopped-flow spectrometers (k_{obs} typically > 0.03 s^{-1}), or a Cary Eclipse fluorescence spectrophotometer (fluorimeter) (Varian) (k_{obs} typically < 0.03 s^{-1}). A pre-formed complex of peptide:partner was mixed with various concentrations of a peptide competitor. To extract the k_{obs}, the change in fluorescence upon formation of the new equilibrium was monitored, averaged (mean) and fit to a single exponential decay function (Equation 2.25). The determined k_{obs} depended on the ratio of competitor to complex ratio: at low ratios, the new equilibrium can contain some peptide:partner complex, and k_{obs} represents the k_{off} plus a component of k_{on}. Whereas, at high concentration ratios of competitor, there is effectively no peptide:partner complex reformation and k_{obs} represents k_{off}.

3.14.1 MLL:KIX and c-MYB:KIX

Temperatures were maintained at 10°C for c–MYB:KIX and 5°C for MLL:KIX. The lower temperatures were required to follow the fast observed rates of c-Myb:KIX and MLL:KIX. Excitation wavelengths of 493 nm and 594 nm were used in conjunction with 515 and 615 nm long pass filters for FITC-MLL and Alexa594-c-MYB labelled peptides, respectively. Dye-labelled peptide (0.25 – 0.5 μM) was pre-equilibrated with 1 – 2.5 μM of KIX and mixed in a 1:1 volume ratio with various concentrations (0 – 50 μM) of unlabeled c-MYB. Unlabelled c-Myb contained a C50A mutation (C322 in the full-length protein) to prevent disulphide bond formation. Data within the dead-time of mixing (the first 1 ms) were removed before fitting.

To extract the k_{off}, the concentration of competitor was plotted against k_{obs} and fitted to Equation 3.9:

$$k_{obs} = k_{off} + k_{on}[\text{partner}] \frac{1}{1+\frac{[\text{competitor}]}{K_{d(\text{competitor})}}} \quad (3.9)$$

where [competitor] is the concentration of competing peptide and $K_{d(competitor)}$ is the equilibrium dissociation constant for the interaction between the competitor and partner protein.

3.14.2 BCL-2 Family

Temperature was maintained at 25°C. Depending on the fluorophore, excitation/emission wavelengths of 340/360 or 320/360 nm (intrinsic tryptophan), and 555/575 nm (TAMRA) were utilised. For reactions monitored on the fluorimeter, excitation and emission slit widths were typically 2.5 nm and 10 nm, respectively. A 320 nm long-pass filter was employed in stopped-flow. For some PUMA-TAMRA mutants, changes in fluorescence intensity were small. Binding kinetics for these mutants were consequently followed by monitoring the change in fluorescence anisotropy, or the change in fluorescence intensity when exciting with and detecting vertically polarised light: the fluorimeter did not have the ability to monitor two fluorescence channels simultaneously. Therefore, fluorescence was monitored only in the VV channel and anisotropy was not calculated.

Peptides (1 – 5 μM) were pre-equilibrated with equimolar amounts of partner and mixed with various concentrations of unlabelled PUMA 35 WT or PUMA W133F/N149A. PUMA W133F/N149A was required for reactions that were following changes in intrinsic tryptophan fluorescence: this mutant had lower intrinsic fluorescence (W133F) and contained a mutation that was shown by Dr Joe Rogers to have a negligible fluorescence intensity change upon binding MCL-1 (Rogers *et al.*, 2014b). For mixing, a 1:9 or 1:10 volume ratio of complex to competitor was used for the fluorimeter and stopped-flow, respectively. For reactions monitored on the fluorimeter, samples were manually mixed before returning to the fluorimeter, which had been left recording. Consequently, the start of the reaction was missed.

To account for this, the time between adding the competitor and returning to the fluorimeter was recorded, and the reaction timescale adjusted accordingly. At least three points, where no dependence of concentration of competitor on k_{obs} was observed, were averaged (mean) to ascertain the k_{off}: when a sufficiently high concentration of competitor is used, the k_{obs} becomes equal to k_{off}.

3.15 Fitting of MLL P9/21A and L8A Binding Kinetics

For reasons described in Chapter 4, the observed binding rates for MLL were fast and displayed small amplitudes. Under these conditions, it proved challenging to extract consistent rates from the mean trace. Consequently, Dr Sarah Shammas and I developed a different fitting method.

Between 60 and 120 traces were collected for each concentration of excess partner protein/peptide. Each trace was individually fit to a single-exponential decay function (Equation 2.25). The individual rate constants obtained for single traces (at each concentration of excess partner protein) were plotted as histograms and fit to a Gaussian function (Equation 3.10) to extract an average rate:

$$f(x) = \frac{a \exp\left[-\frac{(x-\mu)^2}{2\sigma}\right]}{\sigma\sqrt{2\pi}} \tag{3.10}$$

where a is a scaling constant, μ is the mean, and σ is the standard deviation.
To assist Gaussian fitting, only individual fitted rates between 0 and 1000 s^{-1} were included: rates below 0 s^{-1} have no physical meaning, and rates above 1000 s^{-1} cannot be feasibly detected with the stopped-flow setup used to collect these data. To uncover and minimize any potential dependence of the extracted rate on the chosen bin size, 10 bin sizes (from 5 to 50 in increments of 5) were used to generate histograms. k_{obs} was determined as the mean μ over all bin sizes. k_{on} was obtained from the gradient of the straight line fit of k_{obs} versus partner protein concentration.

Dissociation k_{obs} values at each concentration of excess competing peptide were determined using the Gaussian fitting method described above. k_{off} was then obtained by taking the mean of dissociation k_{obs} at different concentrations of out-

competitor: at high concentrations of competitor, there is effectively no complex reformation and k_{obs} represents k_{off}.

These described data analyses for MLL P9/20A and L8A were performed using a bespoke script created in Mathematica, written by Dr Sarah Shammas.

3.16 Equilibrium Binding Affinities

Changes in fluorescence anisotropy were monitored using Cary Eclipse fluorescence spectrophotometers (Varian), equipped with a manual polariser accessory. Temperature was maintained at 10°C for c-MYB, 5°C for MLL and 25°C for BCL-2 family experiments. Depending on the fluorophore, excitation/emission wavelengths of 593/625 nm (a594-c-MYB, Chapter 4), 493/515 nm (FITC-MLL, Chapter 4), and 555/575 nm (TAMRA-Peptides, Chapters 7 and 8) were utilised. Excitation and emission slit widths were typically 2.5 nm and 10 nm, respectively.

The fluorescence intensity in vertically and horizontally polarised channels is used to calculate the anisotropy. Therefore, to take into account the ability of the fluorimeter to detect vertically and horizontally polarised light, a correction factor has to be included. This is known as the *G* factor, and was calculated using Equation 3.11:

$$G = \frac{I_{\mathrm{HV}}}{I_{\mathrm{HH}}} \tag{3.11}$$

where I_{HH} is the fluorescence intensity when using horizontal excitation and emission polarisers, and I_{HV} is the fluorescence intensity when using horizontal excitation and vertical emission polarisers.

Accounting for the *G* factor, the anisotropy (*R*) can then be calculated from Equation 3.12:

$$R = \frac{I_{\mathrm{VV}} - GI_{\mathrm{VH}}}{I_{\mathrm{VV}} + 2GI_{\mathrm{VH}}} \tag{3.12}$$

where I_{VV} is the fluorescence intensity when using vertical excitation and emission polarisers, and I_{VH} is the fluorescence intensity when using vertical excitation and horizontal emission polarisers.

When dye-labelled peptide binds to its partner protein, the rate of tumbling can change, due to the change in the relative size of the peptide. As the rate of tumbling decreases, there is an increase in the amount of light emitted in the plane that was used to excite of the dye (i.e. I_{vv}). However, a change in the fluorescence intensity, as well as the anisotropy, can also observed upon binding. In these cases, the observed changes in I_{VV} are composites of the change in quantum yield/quenching of the dye and the change in the tumbling. This can result in deviations between the calculated anisotropy and the true anisotropy, which influence the determined binding affinity. Dandliker *et al* provide a method to correct for the change in fluorescence intensity upon binding (Dandliker *et al.*, 1981), which can be computed using Equation 3.13:

$$R_{\mathrm{adj}} = \frac{\frac{x}{y} \times \frac{Q_{\mathrm{f}}}{Q_{\mathrm{b}}} \times R_{\mathrm{b}} + R_{\mathrm{f}}}{1 + \frac{x}{y} \times \frac{Q_{\mathrm{f}}}{Q_{\mathrm{b}}}} \tag{3.13}$$

where R_{adj} is the adjusted anisotropy, *Q* is the measured fluorescence intensity, *R* is the measured anisotropy, *x* is the difference between the measured anisotropy and anisotropy of the free peptide, and *y* is the difference between the anisotropy of the bound peptide and the measured anisotropy. The subscripts f and b indicate

that the value is measured for the free or bound peptide, respectively. The total fluorescence intensity of the free (Q_f) and bound (Q_b) peptide was calculated using Equation 3.14:

$$Q = I_{VV} + 2I_{VH} \tag{3.14}$$

3.16.1 Saturation Binding

Dye-labelled peptides were incubated with varying concentrations of partner protein, protected from light, for at least 10 half-lives (calculated from the k_{off}). Samples were transferred to a temperature equilibrated fluorescence cuvette and placed into a temperature controlled sample-handling unit. The fluorescence intensity in each plane was calculated from the mean of three repeats. After calculation of the anisotropy, the equilibrium binding dissociation constant (K_d) was determined using Equation 3.15:

$$R_{adj} = R_f + \Delta R \left(\frac{K_d + [A] + [B] - \sqrt{((K_d + [A] + [B])^2 - 4[A][B])}}{2[A]} \right) \tag{3.15}$$

where [A] is the concentration of dye-labelled peptide, [B] is the concentration of partner protein, R_f is the anisotropy of the free peptide and ΔR is the difference between the free and bound peptide anisotropies.

3.16.2 Competition Binding

PUMA-TAMRA A139G A150G (300 nM) was incubated with 300 nM of MCL-1 and various concentrations of competing peptides. The PUMA A139G A150G mutant was chosen as the K_d could be independently determined from both kinetics and saturation binding experiments. The fluorescence intensity in each plane was calculated from the mean of three repeats. To obtain the midpoint of the binding isotherm, the total fluorescence, calculated using Equation 3.14, was fit to a four-parameter logistic Hill equation (Equation 3.16):

$$Q = Q_{\mathrm{f}} + \frac{\Delta Q}{\left(1+\left(\frac{[\mathrm{Competitor}]}{\mathrm{IC50}}\right)^{-\mathrm{HillSlope}}\right)} \quad (3.16)$$

where IC50 is the midpoint of the binding isotherm, [competitor] is the concentration of competing peptide and HillSlope is the steepness of the curve.

If the K_d of the labelled peptide is known, then the IC50 can be converted to a relative binding affinity (K_i). K_i values for the competing peptides were calculated using the method of Nikolovska-Coleska *et al.* (Nikolovska-Coleska *et al.*, 2004), which is summarised in Equation 3.17:

$$K_i = \frac{I_{50}}{\frac{L_{50}}{\frac{K_{\mathrm{d}}+P_0}{K_{\mathrm{d}}+1}}} \quad (3.17)$$

where I_{50} is the concentration of free inhibitor at 50% inhibition, L_{50} is the concentration of the free labelled ligand at 50% inhibition, P_0 is the concentration of free protein when no competitor is present, and K_d is the dissociation constant for the Labelled-peptide:partner complex.

3.17 Equilibrium Folding Stability

Dilutions of the chemical denaturants, urea or guanidine hydrochloride (GdmCl), were produced using a Microlab 500 liquid processor (Hamilton Company). Protein was incubated at 25°C with the chemical denaturant for at least 2 hours. For A1 and BCL-XL, 10 mM DTT was included, to prevent the formation of disulphide bonds. Protein unfolding was monitored either by following the change in intrinsic tryptophan/tyrosine fluorescence (LS55 luminescence spectrometer - Perkin Elmer, or Cary Eclipse spectrophotometer - Varian), or by monitoring the change in CD mdeg signal at 222 nm (Applied Photophysics Chirascan). Samples were transferred to a 10 mm by 4 mm cuvette, utilising the 10 mm path length for excitation in fluorescence, and the 4 mm path length for absorbance in CD. For fluorescence, a scan from 300 – 400 nm was performed with an excitation

wavelength of 280 nm and slit widths of 10 nm. Settings for CD were typically 2 nm bandwidth, 15 s trace with 500 points.

3.18 Data Analysis

Data were typically fit using Kaleidagraph (version 4.1, Synergy). Analysis of MLL P9/21A and L8A binding kinetics was achieved using Mathematica (Wolfram). Association kinetics for PUMA shuffles and core swaps were analysed using ProFit (version 6.2.16, Quantum Soft).

Chapter 4

Conserved Helix-Flanking Prolines in IDPs

The script used to analyse MLL P9/21A and MLL L8A was written by Dr Sarah Shammas. Dr Gary Daughdrill and Dr Wade Borcherds performed NMR experiments. Dr Sarah Shammas performed and analysed p53 kinetic experiments.

4.1 Structural Features of Prolines

Being the only imino acid, proline is unique amongst the residues that make up protein sequences. Instead of the usual bonds to 2 carbons, and a hydrogen, the backbone nitrogen of proline is bonded to 2 alkyl carbons and the carbonyl from the previous amino acid, forming a cyclic structure (Figure 4.1). This lack of a backbone amide hydrogen prevents proline from forming hydrogen bonds that are required to stabilise both α-helical and β-sheet secondary structures. Additional effects of the rigid, bulky pyrrolidine ring influence both the phi and psi angles of proline and the backbone geometry of the preceding residue (MacArthur and Thornton, 1991). Thus, both prolines and residues preceding prolines, exhibit atypical Ramachandran plots (MacArthur and Thornton, 1991; Ramachandran *et al.*, 1963; Theillet *et al.*, 2014). Compared to the other 19 naturally occurring amino acids, these distinct structural features of proline typically result in destabilisation of protein secondary structure.

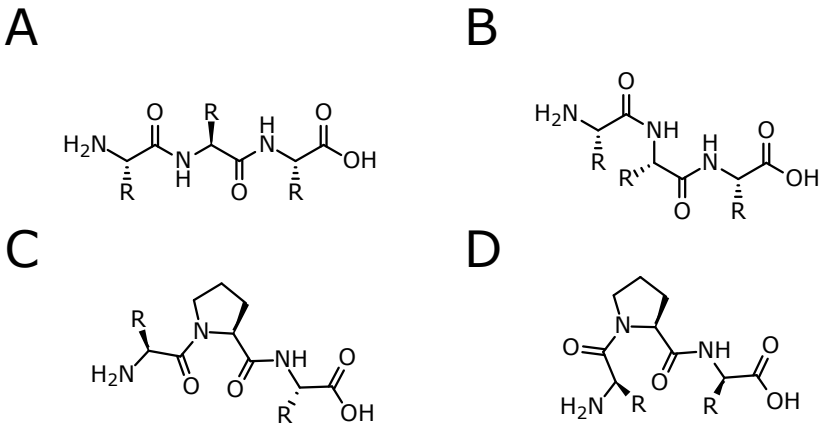


Figure 4.1 Chemical Structure of Amino and Imino Acids in *Cis* and *Trans*. More than 99% of amino acids occur in *trans* (A) rather than *cis* (B). Proline has a smaller difference in free energy between the *tran*s (C) and *cis* (D) forms, resulting in up to 20% occurring the *cis* conformation (Joseph *et al.*, 2012; Theillet *et al.*, 2014). This has an important influence on the local structure: *trans* isoforms result in extended conformations, whereas *cis* isoforms favour turn-like structures.

4.2 Position of Prolines

Due to its destabilising effects, prolines are relatively rarely found in the middle of α-helicies or β-sheets. However, when prolines are located in these positions, they can confer specific functions. For example, the presence of prolines in helical transmembrane segments of membrane channels, transporters and receptors has been observed to occur with relatively high frequency (Deber *et al.*, 1986). While the exact physiochemical reasons for the presence of all these prolines are not fully understood, they have been shown to be crucial for the correct voltage gating of the connexin 26 channel (Suchyna *et al.*, 1993). Furthermore, it has been suggested that the ability of prolines to undergo *cis*/*trans* isomerisation could be a means of regulating channel opening and closing (Williams and Deber, 1991).

Inclusion of proline in the middle of an α-helix can introduce a kink of approximately 26° (Barlow and Thornton, 1988), or can terminate the helical segment by creating a tight turn (Nilsson and von Heijne, 1998). Indeed, in folded proteins, a preference for prolines in positions from the N-cap to N-cap$_{+2}$ and C-cap to C-cap$_{+3}$ has been noted (Richardson and Richardson, 1988; Theillet *et al.*, 2014). In IDPs, proline residues that flank regions of partially formed residual structure, were found to occur more frequently than elsewhere within the IDP (Fuxreiter *et al.*, 2007).

Prolines were consequently proposed to be conserved in flanking regions to 'preserve levels of disorder in regions with residual structure propensities' (Theillet *et al.*, 2014). Combining this with the view that the level of residual 'bound-like' structure determines the binding mechanism (Brown *et al.*, 2011; Krieger *et al.*, 2014), helix-flanking prolines were proposed to act as '"switches" that control the level of helical content' and 'govern the degree of conformational selection by target proteins' (Lee *et al.*, 2014). However, kinetic information is required to determine mechanism (Gianni *et al.*, 2014; Shammas *et al.*, 2016).

4.3 Conservation of Helix-Flanking Prolines

Given its reputation as a helix-breaker, it is easy to imagine that to define helical segments, evolution has resulted in prolines maintaining flanking positions. Figure 4.2 demonstrates the conservation of helix-flanking prolines in six IDPs that undergo coupled folding and binding to form α-helicies.

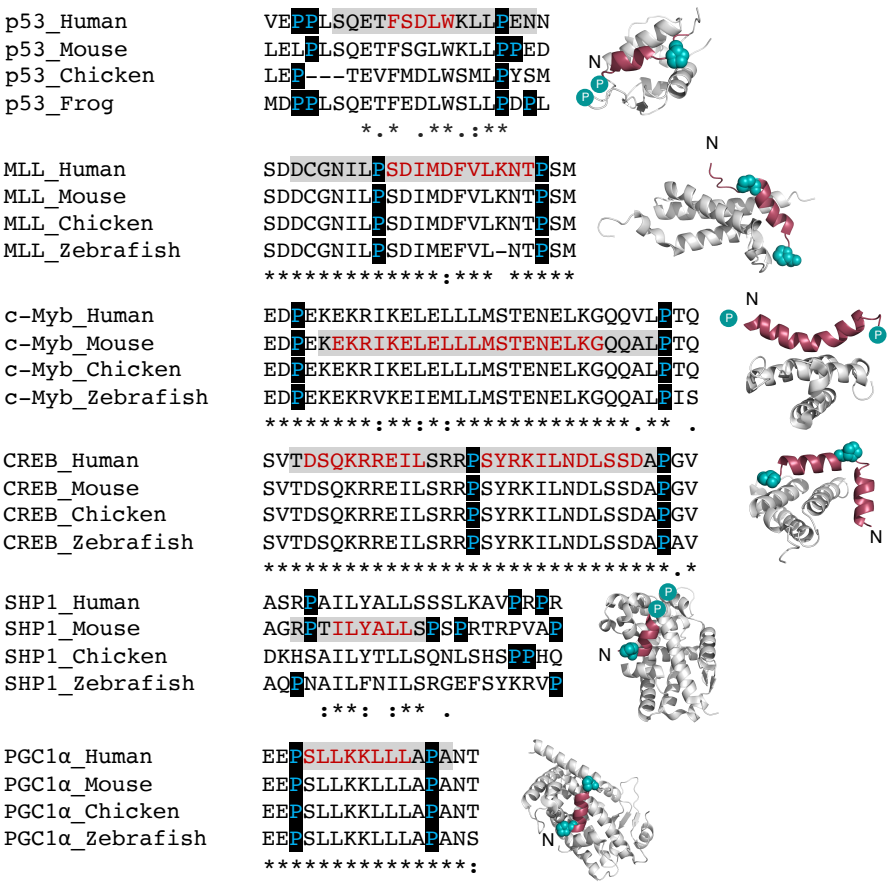


Figure 4.2 Position and Conservation of Helix-Flanking Prolines in IDRs. Bound structures of IDPs/IDRs, which undergo coupled folding and binding to form α-helices that are flanked by prolines. From top to bottom: p53:MDM2 (PDB:1YCR); MLL:CBP KIX (PDB:2LXS); c-Myb:CBP KIX (PDB:1SB0); CREB:CBP KIX (PDB:1KDX); SHP-1:SF-1 (PDB:1YMT); PGC-1 α:PPARγ (PDB:3CS8). Folded partner proteins are shown in grey, IDPs/IDRs in red. Helix flanking prolines are indicated as cyan spheres. Cyan circles indicate helix-flanking prolines that are not present in the structure. N denotes the N-terminus of the IDP/IDR. For each IDP/IDR, the sequence from *Homo sapiens*, *Mus musculus*, *Gallus gallus* and *Danio rerio* was aligned using Clutsal Omega (Goujon *et al.*, 2010; Sievers *et al.*, 2014), p53 alignment is from Borcherds *et al.* (2014). The peptide region used in the structure is highlighted in grey, with residues that form α-helices upon binding in red. Helix-flanking prolines are highlighted. As determined by Clustal Omega, positions of full residue conservation are indicated by an asterisk, a colon indicates conservation of strongly similar amino acid properties, weakly similar properties are specified with a period.

4.4 Helix-Flanking Prolines in p53

To investigate the role of conserved helix-flanking prolines, Borcherds *et al.* investigated proline to alanine mutants in p53. In terms of helix stability, alanine is at the other end of the spectrum to proline (Horovitz *et al.*, 1992; O'Neil and DeGrado, 1990). Therefore, if proline is acting to reduce helix stability, mutation to alanine should provide a large change in observed helicity. Consistent with the idea that conserved helix-flanking prolines control levels of residual structure, mutation of the C-terminal helix-flanking proline of p53 to alanine causes a large increase in helicity (Borcherds *et al.*, 2014). The 2.5-fold gain in helicity occurs within the region that becomes helical upon binding to the structured protein, MDM2 (Figure 4.3). Mutation of P12/13 to alanine also increases helicity, however, to a far lesser extent (32% peak helicity *vs* 28% in WT).

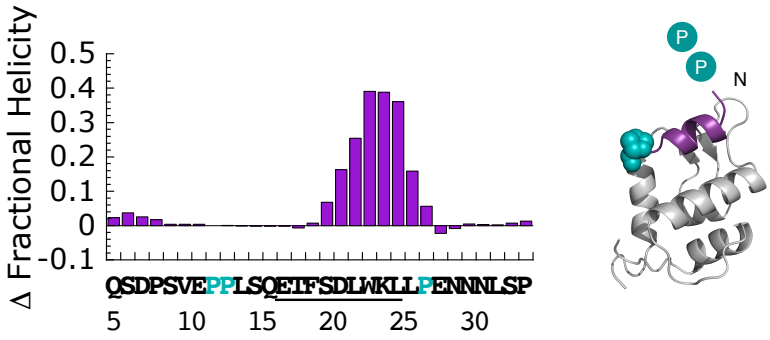


Figure 4.3 Helix-Flanking Prolines Modulate Residual Structure in p53. The data represent the difference in per residue fractional helicity between WT p53 and p53 P27A (Borcherds *et al.*, 2014). Mutation of P27 to alanine increases the helicity of p53 within the region that becomes helical upon binding MDM2. Peak helicity increases by approximately 40%, while the total helicity across all the residues that become helical upon binding increases from 12% in WT p53 to 30% in p53 P27A. Helix flanking prolines are highlighted in cyan in both the sequence and p53:MDM2 structure (PDB entry 1YCR, p53 in purple, MDM2 in grey). As determined from the bound structure, the residues that become helical upon binding MDM2 are underlined. Helix-flanking prolines that fall outside of the peptide used in the p53:MDM2 structure (residues 15 – 29) are indicated as cyan spheres. N denotes the N-terminus of p53.

In terms of the thermodynamics of the p53:MDM2 interaction, the p53 P27A mutant has an approximate 10-fold enhancement in binding affinity compared to WT p53 (K_d = 25 ± 3 nM *vs* 240 ± 60 nM). Mechanistically, this could either be due to an

increase in the rate of association, or a decrease in the rate of dissociation, or a combination of the two. If the prolines were acting to govern the degree of conformational selection (Lee *et al.*, 2014), then one might predict that an increase in k_{on} is responsible for the enhanced affinity *i.e.* when the P27A mutation is present, a greater amount p53 is in a 'bound-like' conformation, resulting in an increased number of successful collisions and a faster rate of association. To assess the kinetic effect of the P27A mutation, Dr Sarah Shammas performed stopped-flow experiments (Figure 4.4). Despite the 2.5-fold increase in residual 'bound-like' structure, k_{on} is increased by only 1.23 ± 0.06-fold for the P27A mutant. Clearly, an enhanced rate of complex formation is not responsible for the 10-fold gain in affinity of p53 P27A for MDM2. Instead, the tighter binding is explained by a 12-fold decrease in k_{off} (Crabtree *et al.*, 2017).

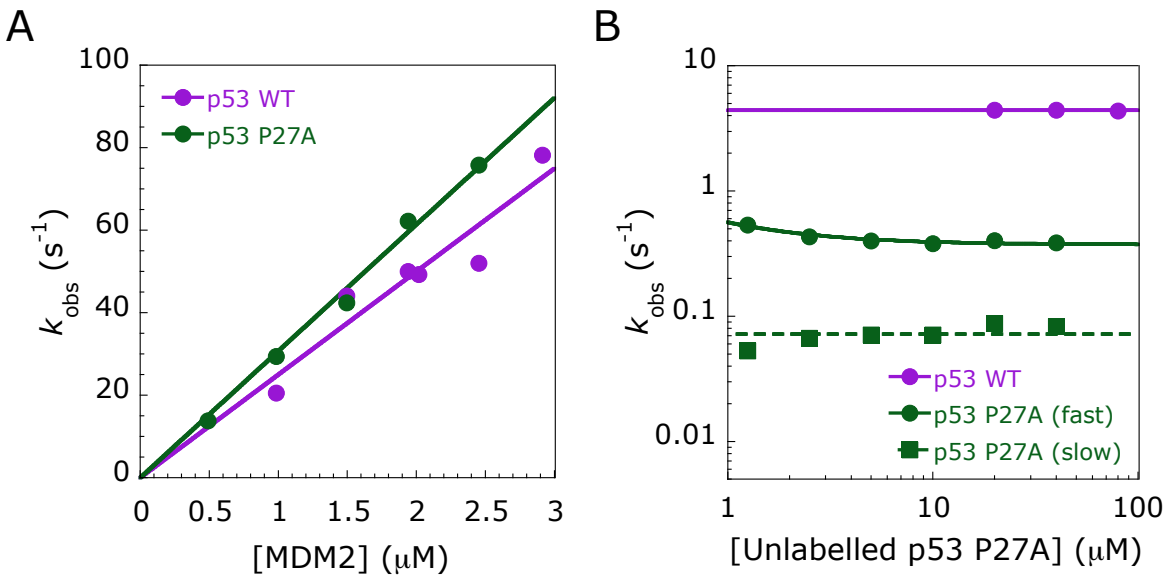


Figure 4.4 P27A Enhances p53:MDM2 Affinity by Affecting k_{off}, not k_{on}. (A) Pseudo-first order association kinetics for p53 binding to MDM2. The gradient of the straight-line fit gives association rate constants of 25 ± 1 $\mu M^{-1}s^{-1}$ for WT and 30.7 ± 0.7 $\mu M^{-1}s^{-1}$ for P27A. (B) Observed dissociation rates for p53:MDM2, obtained from out competition experiments. The k_{off} is given by the asymptote of the fit. There are 2 dissociation rate constants for the p53 P27A mutant: a 'fast' rate of 0.37 ± 0.01 s^{-1} and a 'slow' rate of 0.072 ± 0.005 s^{-1}. WT p53 has a single k_{off} of 4.40 ± 0.03 s^{-1}.

In summary, the conserved helix-flanking prolines in p53 act to restrict the level of residual helicity. However, this is not to control the rate of association through modulating the level of binding competent p53. Instead, the change in k_{off} demonstrates that the influence of the altered helix stability is on the bound complex. To determine if this was a common mechanism of action, helix-flanking prolines in two further IDRs, the transactivation domains of MLL and c-MYB, were investigated.

4.5 Results

Upon binding to distinct sites on CBP-KIX, both MLL and c-MYB fold and bind to form α-helices that are flanked by prolines. (Figure 4.2). Similar to the p53 study, the helix-flanking prolines were mutated to alanine. Both the effects on the residual structure and the kinetics and thermodynamics of the interaction with CBP-KIX were then monitored. Single and double proline mutants were investigated for both MLL (P21A, P9/21A) and c-MYB (P17A, P17/44A).

4.5.1 Helix-Flanking Prolines Reduce Residual Helicity in MLL and c-MYB

By measuring the difference in a protein's absorbance of right and left handed polarised light, circular dichroism can provide structural information. WT and mutant peptides were scanned to determine the overall change in residual helicity upon helix-flanking proline to alanine mutation. Comparison of the WT peptides demonstrated that MLL was mostly disordered (13% estimated helicity), whilst c-MYB contained some residual helicity (25% estimated helicity) (Figure 4.5). Mutation of the C-terminal helix flanking proline in MLL (P21A) results in a small increase in helicity (15% estimated helicity). The double mutant (P9/21A) demonstrates a further enhancement in helical content (17% estimated helicity). Whilst these absolute changes are small, relative to the largely disordered WT peptide, the P9/21A mutant represents a 1.3-fold increase in overall helical content compared to WT. A similar 1.3-fold change in overall helicity was observed for the c-MYB P17A mutant (32% estimated helicity). Mutation of both c-MYB helix-flanking prolines resulted in a 1.4-fold increase in helicity (35% estimated helicity) compared to WT. As the signal from CD is an average across all the peptide bonds, these reported changes in helicity represent the change across the entire peptide. A residue-level technique is required to locate the changes in helicity.

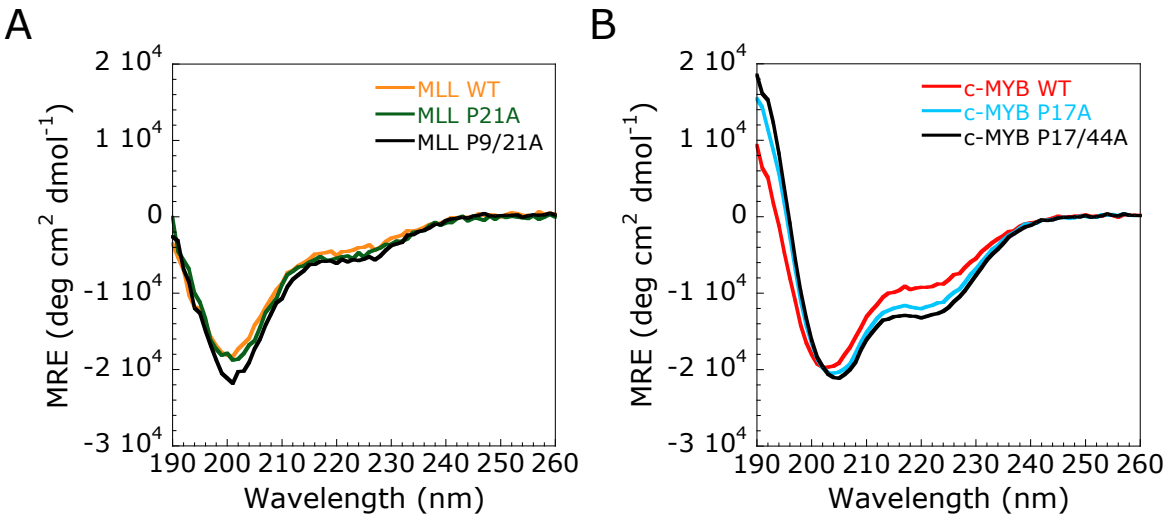


Figure 4.5 Overall Residual Helicity in MLL and c-MYB is Increased upon Proline to Alanine Mutation.
CD data for (A) MLL and (B) c-MYB peptides. A decrease in the MRE at 222 nm is indicative of an increase in α-helical content. Data are the mean of three independent scans.

To ascertain whether the increases in helicity occurred within the region that becomes helical upon binding, NMR was utilised – NMR can provide residue-specific chemical shifts, which can be entered into the δ2D algorithm (Camilloni *et al.*, 2012) to produce residue-level helicity estimates. The double helix-flanking proline to alanine mutants demonstrated the biggest changes in residual structure in the CD experiments. Therefore, both WT and the double mutant peptides (MLL P9/21A and c-MYB P17/44A) were chosen for further investigation by NMR.

Whilst the NMR-based estimated helicity for c-MYB WT (22% *vs* 25% from CD) and P17/44A (29% *vs* 35% from CD) matched the CD data reasonably well, a substantially lower NMR–based helicity for MLL WT (2% *vs* 13% from CD) and P9/21A (3.5% *vs* 17% from CD) was observed (Figure 4.6).

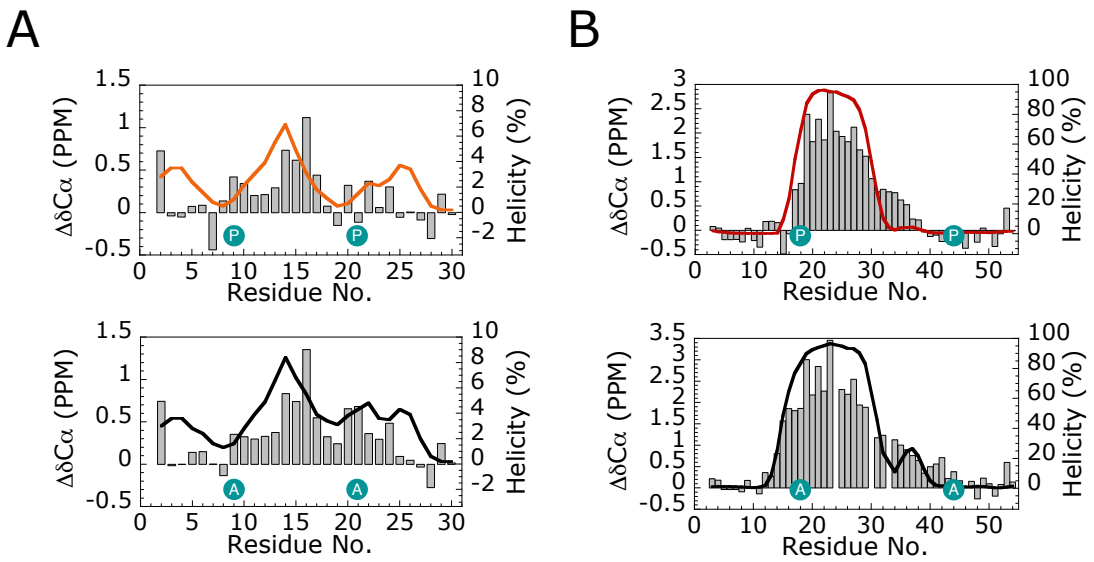


Figure 4.6 Chemical Shift and Per Residue Helicity for MLL and c-MYB. Per-residue helicity for (A) MLL WT (top), MLL P9/21A (bottom), (B) c-MYB WT (top) and c-MYB P17/44A (bottom). Chemical shifts were determined from ^{1}H-^{15}N HSQC and three-dimensional HNCACB and HNCO experiments. To calculate secondary Cα chemical shifts (bars), residue-specific random coil chemical shift values were obtained from the ncIDP library (Tamiola *et al.*, 2010) and subtracted from the measured chemical shift. Percentage helicity (line) was calculated from Cα, CO, Cβ, HN and N secondary chemical shifts using the δ2D algorithm (Camilloni *et al.*, 2012). Cyan spheres represent the position of WT proline, or mutant alanine residues. Data were collected at 10°C. For c-MYB P17/44A, the chemical shift for 3 residues could not be observed. This was likely due to the rate of chemical exchange at 10°C, as the chemical shifts could be detected at 25°C. Overall helicity was similar at 25°C and 10°C for c-MYB WT (23% *vs* 22%) and c-MYB P17/44A (29% *vs* 28%).

There are a few potential explanations for the discrepancy between CD and NMR for MLL. Firstly, the Cα and CO secondary chemical shifts did not match (Figure 4.7). The Cα shifts were typically positive, indicating that the residues were, at least partially, experiencing a helical environment. However, the CO shifts were generally lower than those observed for Cα, and the sign of the shift was opposite for some residues, especially in WT MLL, indicating a less helical environment. Due to the sensitivity of the carbonyl to hydrogen bonding and solvent exposure, prediction of helicity from CO secondary chemical shifts can be less reliable than using Cα shifts (Wishart and Sykes, 1994). To calculate secondary structure content, the δ2D algorithm (Camilloni *et al.*, 2012) utilised Cα, CO, Cβ, HN and N secondary chemical shifts. Consequently, the use of the inconsistent CO shifts could

have resulted in an underestimation of the helicity. Secondly, whilst the trends are similar, the absolute helicity determined from CD is dependent on the method chosen. For example, using the method of Chen *et al.* the estimated helicity for WT MLL is 7%, compared to 13% estimated using the method of Muñoz and Serrano (Chen *et al.*, 1974; Muñoz and Serrano, 1995). Nevertheless, despite the discrepancy in absolute helicity, the total fold change in helicity across the entire peptide for MLL (1.5-fold *vs* 1.3-fold) and c-MYB (1.3-fold *vs* 1.4-fold) was in reasonable agreement with CD data.

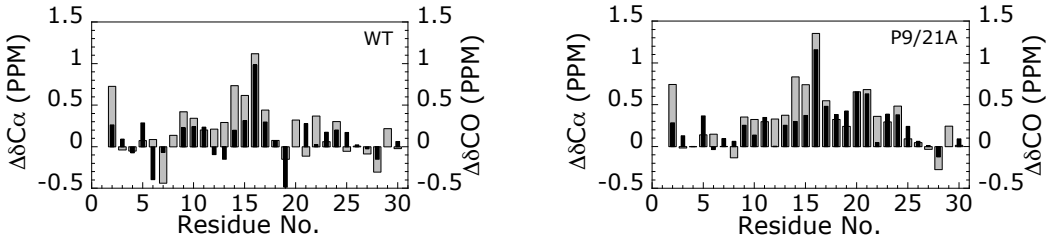


Figure 4.7 Comparison of MLL Cα and CO Secondary Chemical Shifts.
Cα secondary chemical shifts (grey) were generally higher than CO shifts (black). Positive Cα and CO secondary shifts indicate helical secondary structure. A negative shift indicates β-sheet structure. The position of WT proline, or mutant alanine residues are indicated by cyan spheres.

Similar to p53, the majority of the increase in helicity was observed around the site of the proline to alanine mutation (Figure 4.8). Within the region that becomes helical upon binding, a 1.5-fold increase in helicity was observed for MLL P9/21A. c-MYB P17/44A was increased by 1.2-fold.

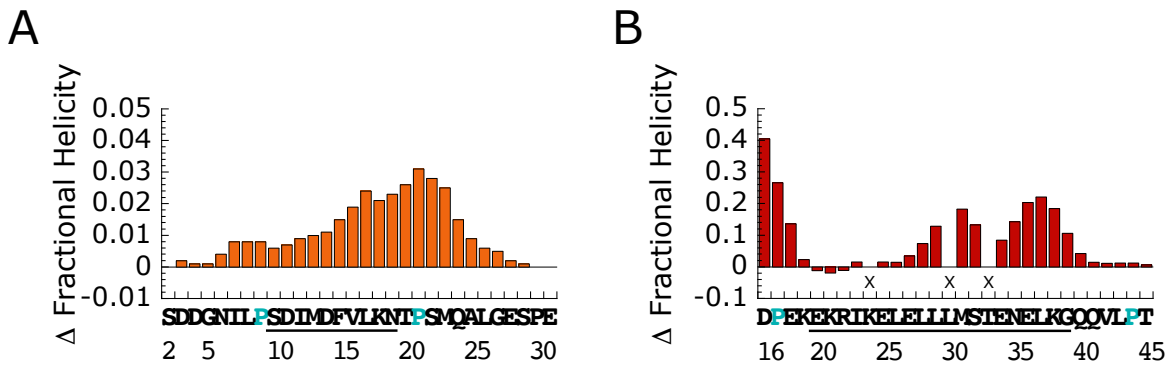


Figure 4.8 Increase in Helicity Occurs within the Region that Becomes Helical upon Binding.

The change in fractional helicity for (A) MLL and (B) c-MYB upon helix-flanking proline to alanine mutation. The change in fraction helicity was calculated by subtracting the WT δ2D helicity estimate from the double mutant δ2D helicity estimates. For both MLL and c-MYB, 30 residues surrounding the region that becomes helical upon binding (underlined) are shown. The position of helix-flanking prolines is highlighted in the amino acid sequence. Residues where the chemical shift value could not be determined at 10°C are indicated with an x.

4.5.2 The Effect of Helix-Flanking Prolines on Binding Affinity is Protein Specific

An increase in residual p53 structure is correlated with an enhancement in binding affinity for MDM2 (Borcherds *et al.*, 2014). Equilibrium binding experiments were employed to establish if the enhanced 'bound-like' structure of MLL and c-MYB also results in an increased affinity for CBP-KIX. In contrast to p53, both MLL and c-MYB demonstrated reductions in binding affinity for CBP-KIX (Figure 4.9). The equilibrium dissociation constant was shifted by 1.9 ± 0.4-fold for c-MYB P17A and 1.6 ± 0.1-fold for P17/44A. Whilst the K_d for the single and double proline to alanine mutants of c-MYB was similar, the difference in the effect of the MLL P21A and P9/21A mutations was stark. Mutation of the C-terminal Helix-flanking proline (P21A) had no effect on the observed K_d (0.7 ± 0.1 μM *vs* 0.66 ± 0.06 μM for WT). Whereas, the additional mutation of the N-terminal helix flanking proline (P9/21A) resulted in a large, 25 ± 4 fold decrease in affinity.

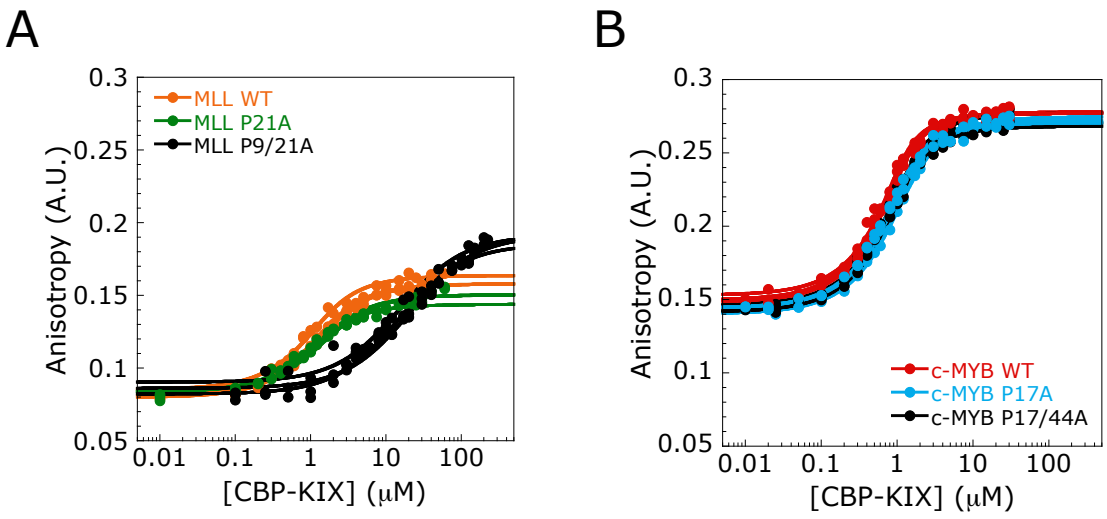


Figure 4.9 Affinity of MLL and c-MYB for CBP-KIX is Reduced by Mutation of Helix-Flanking Prolines to Alanines.
Equilibrium binding curves for (A) MLL and (B) c-MYB binding to CBP-KIX. The binding of fluorescent dye-labelled peptide was followed by monitoring the change in fluorescence anisotropy. Upon binding, the tumbling of the peptide becomes relatively slower, resulting in an increase in anisotropy. Equilibrium dissociation constants of 0.66 ± 0.06 μM (MLL WT), 0.7 ± 0.1 μM (MLL P21A), 17 ± 2 μM (MLL P9/21A), 0.168 ± 0.009 μM (c-MYB WT), 0.32 ± 0.06 μM (c-MYB P17A) and 0.27 ± 0.02 μM (c-MYB P17/44A) were calculated from the mean of three repeats. Errors represent the standard error of the mean.

4.5.3 Reduction in Affinity is Due to an Increase in k_{off}

Stopped-flow kinetic experiments were utilised to determine whether the changes in affinity for MLL and c-MYB were due to alterations in k_{on} or k_{off}.

4.5.3.1 Association Kinetics

The signal changes observed upon mixing c-MYB peptides, or MLL WT and MLL P21A, with KIX were reproducible, fit well to a single exponential and displayed a good signal to noise ratio (Figure 4.10). However, upon mixing the MLL P9/21A mutant with CBP-KIX, the observed signal changes were relatively small, rapid and non-reproducible. This reduction in signal change was at least partly due to the increase in K_d. For example, mixing 0.1 μM MLL WT with 1 μM KIX results in approximately 66% of the peptide going into complex (calculated using Equation 3.3). Whereas, for MLL P9/21A, the reduced affinity results in only 5% going into complex. Consistent with this hypothesis, the amplitude for MLL P9/21A increased when it was mixed with higher concentrations of CBP-KIX. While this improved the

signal to noise, the gain was offset by the faster observed rate constant (k_{obs}), which approached the detection limit of our stopped-flow apparatus. Consequently, it was challenging to extract consistent rates under these conditions of rapid, small signal changes.

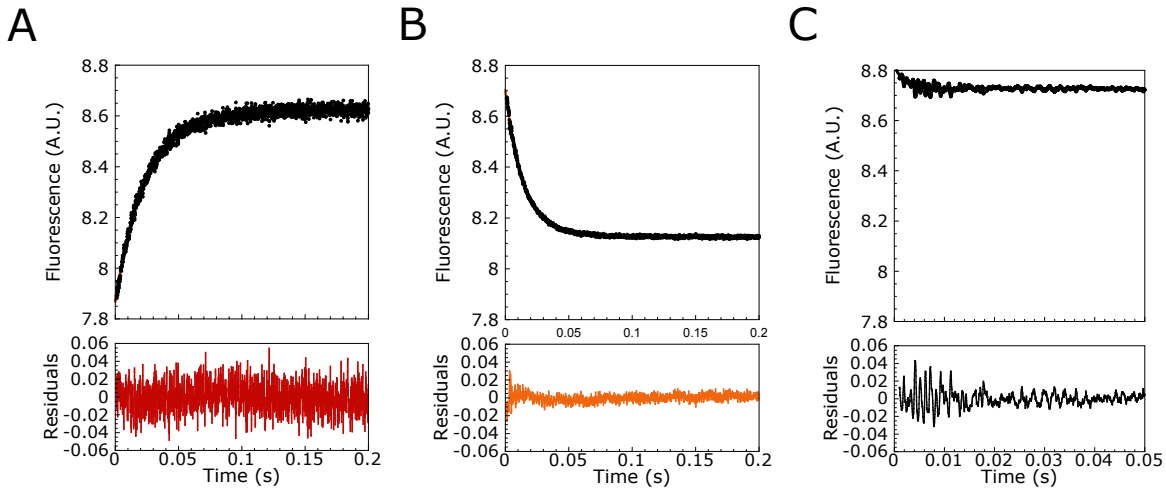


Figure 4.10 Variable Signal Changes for Peptides Binding to CBP-KIX. Example traces for (A) c-MYB peptides, (B) MLL WT and P21A, and (C) MLL P9/21A obtained from fluorescence stopped-flow experiments. Peptide was mixed with an excess concentration of CBP-KIX (A = 1 μM), (B = 1.5 μM), (C = 2.5 μM). Data were collected under pseudo-first order conditions and are fit to a single exponential decay function (Equation 2.25). Traces represent the average of (A) 25, (B) 45 and (C) 94 repeats.

Typically, k_{obs} was determined from the mean of multiple traces. However, the mean for MLL P9/21A became skewed by the inconsistent traces. One way to resolve this issue would have been to manually remove the outlying traces before taking the mean. While convenient, this method introduces significant experimenter bias and was therefore avoided. Instead, a new data analysis method was developed. Individual traces were fit to a single exponential decay function. The extracted k_{obs} were put into bins and fit to a Gaussian function (Equation 3.10), with limits of 0 s^{-1} and 1000 s^{-1}. These upper and lower bounds were chosen as negative rates have no physical meaning and rates over 1000 s^{-1} could not be detected using our stopped-flow apparatus. This method of clustering the fitted data, employing limits and fitting the posterior distribution to a Gaussian function is akin to analysing the data using Baye's theorem. For WT MLL, the distribution of

rates was tight (Figure 4.11), and the mean of the Gaussian function was very similar to the rate obtained using the typical method of extracting k_{obs} from the average trace (e.g. 74.5 s^{-1} *vs* 74.2 s^{-1}). On the other hand, the k_{obs} for MLL P9/21A were widely distributed between 0 and 2000 s^{-1}. The variation and distribution of k_{obs} for MLL P9/21A was dependent on the reaction rate and signal change. Discrepancies between the two fitting methods were, consequently, dependent on the concentration of CBP-KIX (Figure 4.12).

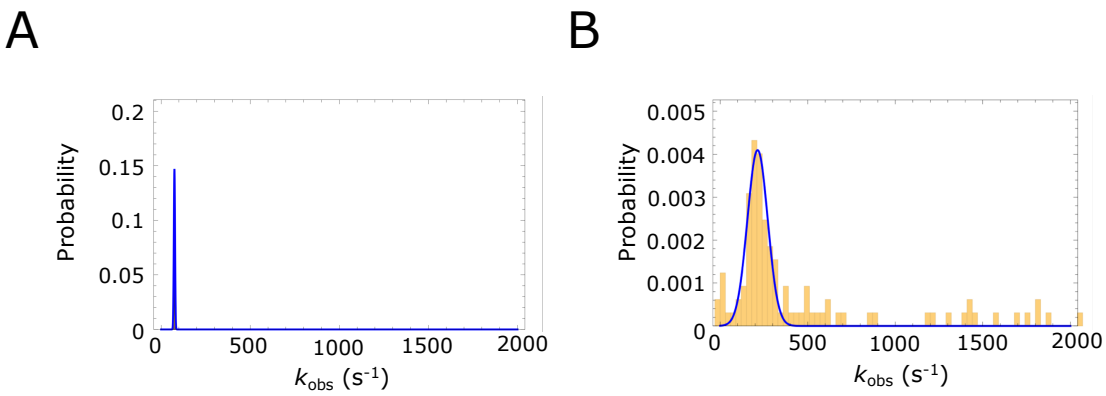


Figure 4.11 Example Histograms for the k_{obs} of MLL WT and P9/21A.
The traces obtained for (A) 2.7 μM CBP-KIX binding to MLL WT, and (B) 2 μM CBP-KIX binding to MLL P9/21A were individually fit to obtain the k_{obs}. The histogram shows the k_{obs} binned into a bin size of 5. There was a large variation in the k_{obs} determined for MLL P9/21A, which skewed the k_{obs} obtained from the mean trace. For example, averaging all of the traces first and then fitting to a single exponential decay function gave a k_{obs} of 310 ± 12 s^{-1}. Whereas, fitting all the traces individually and obtaining k_{obs} from the mean of the Gaussian fit gave a rate of 214 ± 56 s^{-1} (B). This discrepancy between the two methods was dependent on the variation in k_{obs} – for (A) MLL WT, which had a tight distribution, the k_{obs} from each method was similar (74.5 s^{-1} *vs* 74.2 s^{-1}). To overcome any dependence on bin size, 10 bin sizes, from 5 – 50 in increments of 5, were used to produce estimates of k_{obs} from the Gaussian mean. The k_{obs} for the reaction was then determined from the mean of the Gaussian means. The errors in k_{obs} are the standard deviation of the Gaussian fit and therefore represent the distribution of the dataset.

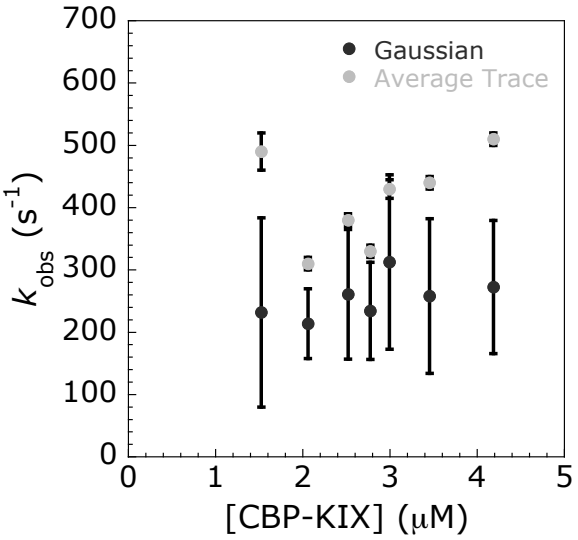


Figure 4.12 k_{obs} for MLL P9/21A Obtained from the Gaussian and Mean Trace Methods.
Error bars represent the curve fit error for the k_{obs} obtained from the average fit method and the standard deviation for the k_{obs} determined using the Gaussian fit method. The errors for the Gaussian k_{obs} therefore indicate the distribution of the dataset.

As k_{obs} is a composite of the k_{on} and k_{off} (Equation 3.8), both rate constants can, in theory, be extracted from the straight line fit of k_{obs} *vs* the concentration of excess protein (CBP-KIX). While this works well for estimating the k_{on}, the determined k_{off} may not be accurate, especially when it is below 1 s^{-1}: a small error in the gradient (k_{on}) will result in a relatively large difference in the intercept (k_{off}). For the c-MYB peptides, no change in k_{on} was observed when either the N-terminal or both helix-flanking prolines were mutated to alanine (Figure 4.13). MLL similarly displayed little change in the k_{on} for either proline mutant. Indeed, the only difference of more than 2-fold for either MLL or c-MYB, was an apparent change in the intercept for MLL P9/21A. The data for this mutant were relatively scattered, due to the rapid observed rates, but the change in intercept indicated an increase of 22 ± 13-fold in the k_{off}.

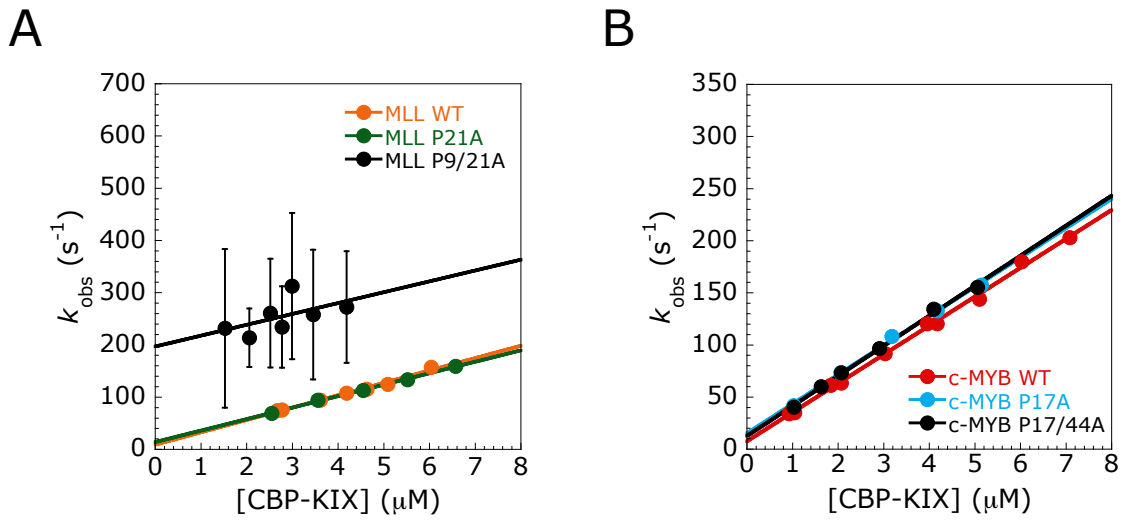


Figure 4.13 Association Rate Constants for c-MYB and MLL Peptides. Observed rate constants for (A) MLL and (B) c-MYB were obtained from stopped-flow experiments. Dye-labelled peptides were rapidly mixed with an excess of CBP-KIX. The change in fluorescence with time was followed and fit to a single exponential to obtain k_{obs}. Error bars for MLL P9/21A represent the standard deviation of the Gaussian fit. The errors for MLL P9/21A therefore indicate the distribution of the dataset. The error of the fit for all other k_{obs} were smaller than the data point. A straight-line fit gives a k_{on} (gradient) of 27.7 ± 0.5 $\mu M^{-1}s^{-1}$ for c-MYB WT, 28.1 ± 0.9 $\mu M^{-1}s^{-1}$ for c-MYB P17A, 28.8 ± 0.8 $\mu M^{-1}s^{-1}$ for c-MYB P17/44A, 24 ±1 $\mu M^{-1}s^{-1}$ for MLL WT, 22.0 ± 0.6 $\mu M^{-1}s^{-1}$ for MLL P21A and 21 ± 14 $\mu M^{-1}s^{-1}$ for MLL P9/21A. Errors in k_{on} represent the error of the fit. The intercept provides an estimate of the k_{off}.

4.5.3.2 Dissociation Kinetics

Association kinetics experiments demonstrated that the helix-flanking proline to alanine mutations had little effect on the k_{on}. A change in the intercept was observed for MLL P9/21A, indicating an increase in $k_{off.}$ To confirm these changes, dissociation kinetics experiments were performed (Figure 4.14). Small fold increases of 1.38 ± 0.02 and 1.41 ± 0.01 were observed for both c-MYB P17A and P17/44A, respectively. As suggested by the association kinetics experiments, a large 15.8 ±0.5-fold increase in k_{off} was determined for MLL P9/21A. The k_{off} for MLL P21A (12.22 ± 0.08) was within error of MLL WT (12.4 ± 0.3).

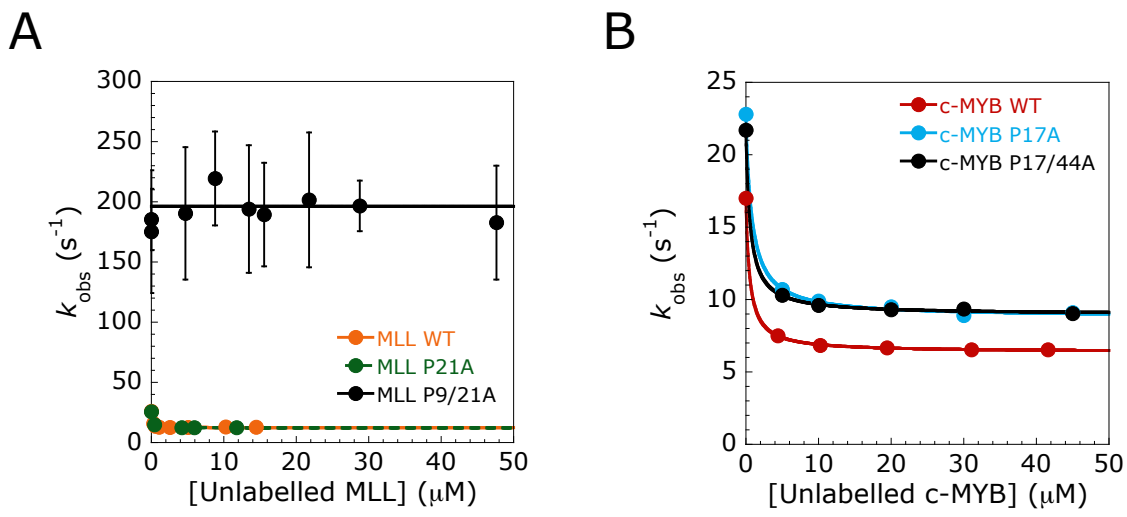


Figure 4.14 Dissociation Rate Constants for MLL and c-MYB Peptides. Observed rate constants for MLL (A) and c-MYB (B) peptides were obtained from stopped-flow experiments. A preformed complex of dye-labelled peptide and CBP-KIX was rapidly mixed with an excess of unlabelled peptide. The change in fluorescence with time was followed and fit to a single exponential to obtain k_{obs}. Error bars for MLL P9/21A represent the standard deviation of the Gaussian fit. The errors for MLL P9/21A therefore indicate the distribution of the dataset. The error of the fit for all other k_{obs} were smaller than the data point. The asymptote gives a k_{off} of 6.38 ± 0.03 s^{-1} for c-MYB WT, 8.8 ± 0.1 s^{-1} for c-MYB P17A, 8.96 ± 0.08 s^{-1} for c-MYB P17/44A, 12.4 ± 0.3 s^{-1} for MLL WT, 12.22 ± 0.08 s^{-1} for MLL P21A and 196 ± 4 s^{-1} for MLL P9/21A. Errors in k_{off} represent the error of the fit, except for MLL P9/21A, where the error represents the SEM (n = 7).

4.5.4 MLL P9A Alters the Interaction of L8 with CBP-KIX, Reducing Binding Affinity

Generally, mutation of helix-flanking prolines to alanines had little effect on the kinetics and thermodynamics of the interaction with CBP-KIX, for either MLL or c-MYB. Mutation of P9A in MLL was the exception to this generality. Despite an apparent increase in residual 'bound-like' structure, P9/21A resulted in a decrease in binding affinity, compared to both MLL WT and P21A. Inspection of the MLL:CBP-KIX structure provided insight into this surprising result (Figure 4.15). A leucine (L8) immediately N-terminal to P9 appeared to pack into a hydrophobic pocket in CBP-KIX in all of the 20 lowest energy NMR structures. CD spectra of the MLL:KIX complex indicated that there was a greater gain in helical structure upon MLL P9/21A interacting with CBP-KIX, compared to the increase observed with WT MLL. Whilst CD is not residue specific and does not provide information on where this

gain in structure occurs, it could be hypothesised that the extra gain in structure is due to extension of the bound MLL helix: prolines are helix-breakers and mutation of the helix-flanking proline could allow extension of the helix. Unfortunately, NMR of the bound complex could not be employed, due to the disappearance of peaks. Another hint towards MLL P9/21A:CBP-KIX having a different bound structure to MLL WT was provided by the equilibrium anisotropy data (Figure 4.9). Anisotropy is related to the tumbling of the complex: the slower the tumbling, the higher the anisotropy. MLL P9/21A displayed a higher maximal anisotropy than either MLL WT or MLL P21A, indicating that the tumbling of the dye was reduced.

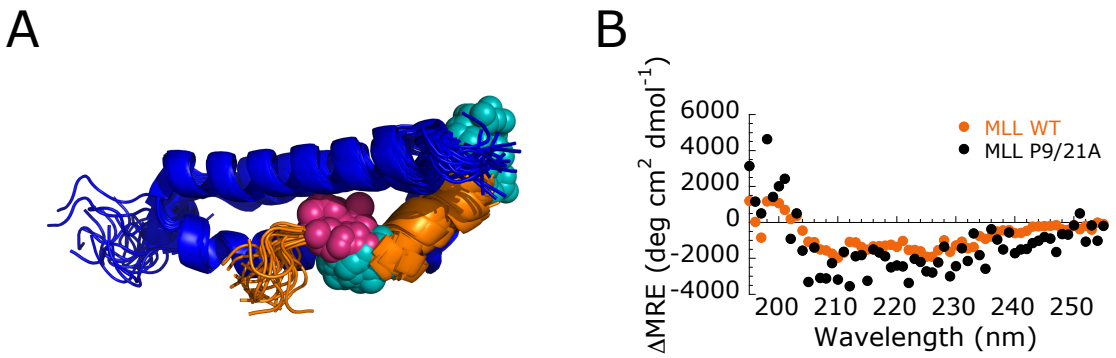


Figure 4.15 Structure of the MLL:CBP-KIX Complex.
(A) The 20 lowest energy NMR structures of the MLL:KIX complex (PDB:2LXS, MLL in orange). Helix-flanking prolines are shown as cyan spheres. L8 is shown as a pink sphere. (B) The difference in MRE between the mean of free KIX and free MLL, and the MRE obtained when the two are mixed, as described in Chapter 3.11.5. Spectra are corrected for the proportion of MLL in complex, therefore indicating the change in CD spectrum if all MLL was bound to KIX.

Extension of the helix in the bound complex could cause the interaction of MLL L8 with CBP-KIX to be disrupted. If this was the case, the removal of L8 by mutation to alanine would be expected to produce similar kinetic and thermodynamic effects to P9A. Mutation of L8A was therefore used as a probe to test the hypothesis that P9A disrupts the interaction of L8 with KIX. A 53 ± 8-fold reduction in affinity was observed for L8A, compared to the 25 ± 4-fold decrease for MLL P9/21A (Figure 4.16). Again, the change in K_d was predominantly explained by an apparent increase in k_{off}.

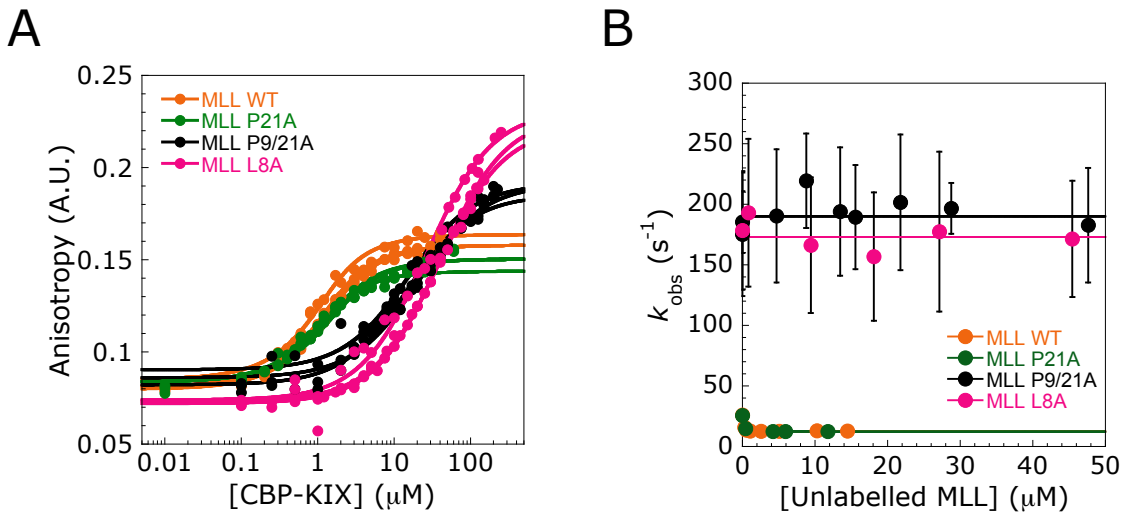


Figure 4.16 L8A Reduces the Affinity of MLL for CBP-KIX by Increasing k_{off}. (A) Equilibrium binding curves for MLL peptides binding to CBP-KIX. The binding of fluorescent dye-labelled peptide was followed by monitoring the change in fluorescence anisotropy. Upon binding, the tumbling of the peptide becomes relatively slower, resulting in an increase in anisotropy. Equilibrium dissociation constants of 0.66 ± 0.06 μM (MLL WT), 0.7 ± 0.1 μM (MLL P21A), 17 ± 2 μM (MLL P9/21A), 35 ± 4 μM (MLL L8A) were calculated from the mean of three repeats. Errors represent the standard error of the mean. (B) Observed rate constants for MLL peptides were obtained from stopped-flow experiments. A preformed complex of dye-labelled peptide and CBP-KIX was rapidly mixed with an excess of unlabelled peptide. The change in fluorescence with time was followed and fit to a single exponential to obtain k_{obs}. Error bars for P9/21A and L8A represent the standard deviation of the Gaussian fit. The errors for MLL P9/21A and L8A therefore indicate the distribution of the dataset. The error of the fit for all other k_{obs} were smaller than the data point. The asymptote gives a k_{off} of 12.4 ± 0.3 s^{-1} for MLL WT, 12.22 ± 0.08 s^{-1} for MLL P21A, 196 ± 4 s^{-1} for MLL P9/21A and 173 ± 6 for MLL L8A. Errors in k_{off} represent the error of the fit, except for MLL P9/21A and L8A, where the error represents the SEM (n = 7 and 4, respectively).

4.6 Discussion

Proline's unique chemical features reduce the stability of protein secondary structure. Whilst they are commonly found throughout IDPs (Ward *et al.*, 2004), their conservation and relative propensity in regions that flank residual helices has sparked interest (Lee *et al.*, 2014; Theillet *et al.*, 2014). In line with the idea that prolines are helix-breakers, mutation of helix-flanking prolines to alanines in p53 (Borcherds *et al.*, 2014), MLL and c-MYB increased the residual helical structure. However, not all proline mutations had equal effects. Mutation of the p53 N-

terminal helix-flanking prolines, P12 and P13, have little effect on the residual helicity compared to the C-terminal P27A mutation, which increases helicity by 2.5-fold (Borcherds *et al.*, 2014). On the other hand, mutation of the N-terminal helix-flanking proline in c-MYB, P17A, increased residual helicity to a greater extent than the C-terminal P44A mutation. MLL had similar increases in helicity when either the C- (P21A) or N-terminal helix-flanking proline was mutated (Table 4.1).

Increasing the proportion of IDP with a 'bound-like' structure has been shown to enhance binding affinity (Borcherds *et al.*, 2014; Iešmantavičius *et al.*, 2014; Krieger *et al.*, 2014). Given that the IDP looks more like the bound state, it is easy to imagine this is due to an enhanced rate of complex formation (Brown *et al.*, 2011; Krieger *et al.*, 2014). However, this was not the case for p53: only a 1.2-fold change in k_{on} occurs, despite a 2.5-fold increase in 'bound-like structure'. Instead, the 10-fold shift in K_d for p53 P27A is predominantly due to a 12-fold decrease in k_{off} (Crabtree *et al.*, 2017). A link between changes in residual structure and dissociation rate constants has also been demonstrated for the IDP PUMA binding to MCL-1 (Rogers *et al*., 2014b), and is consistent with an induced fit mechanism of binding.

Both MLL and c-MYB show increases in residual structure upon helix-flanking proline to alanine mutation. Yet, the affinity of c-MYB P17A, P17/44A and MLL P9/21A for CBP-KIX was reduced. The reduction in affinity for MLL P9/21A was so large that it made following the reaction kinetics extremely difficult, and a new method of data analysis was required to ascertain the k_{on} and k_{off}. While the data were still scattered, especially for the association kinetics, there are reasons to have confidence in the determined rate constants. Firstly, the k_{off} from the intercept of the association kinetics data (200 ± 40 s^{-1}) and from the dissociation experiment (196 ± 4 s^{-1}) were the same. Secondly, the k_{on} calculated by division of the k_{off} by the equilibrium K_d (12 ± 1 $\mu M^{-1}s^{-1}$), was within error of the k_{on} determined from the association kinetics experiment (21 ± 14 $\mu M^{-1}s^{-1}$). The kinetic analysis of c-MYB and MLL peptides demonstrated that, despite observing a 1.2-fold and 1.5-fold increase in 'bound-like' structure, no enhancement in k_{on} occurred. Indeed, the association rate constant for the MLL mutants were, if anything, slightly slower than for WT MLL (Table 4.1). These data are, therefore, inconsistent with the view that helix-flanking prolines act as '"switches" that control the level of helical content' and 'govern the degree of conformational selection by target proteins' (Lee *et al.*,

2014). This conclusion is based on the assumption made throughout this Chapter - that an increase in residual helicity in the region that becomes helical upon binding is equal to an increase in 'bound-like' structure. This assertion was made due to the observation that a single proline mutation increases the average population helicity across a large number of amino acids (both adjacent and up to 13 amino acids away from the site of mutation), suggesting that the entire region is more helical at the same time in the same molecule. This would be achieved through stabilisation of the helical turn where the mutation was made, which subsequently makes the rest of the helix more stable due to the cooperative formation of intra-helical hydrogen bonds. However, while this is the simplest explanation, the methods used to probe residual structure are an average of all the examined molecules. Consequently, it may be that not all of the increase in helicity is present in the same molecule at the same time *i.e.* if residue *i* and *i*+11 have 50% helicity by NMR, *i* may be helical in 50% of the population and unstructured in the rest. *i*+11 could then be helical in the other 50% of the population, with no helicity in the population where *i* is helical. If this was the case, it would mean that the 'bound-like' structure (e.g. a full helix between residues *i* and *i*+11) had not increased, as only a segment of the region would resemble the bound-state in any given molecule. Thus, for a protein that bound via a conformational selection mechanism, there would be no increase in the concentration of binding competent species, resulting in no increase in the k_{on}. Alternatively, the mutation could have reduced the concentration of protein in the 'correct' binding conformation, which would reduce k_{on}. One way to examine this would be to look at the change in structure of a single protein molecule, for example using single molecule FRET or molecular dynamics simulations. Despite this caveat, the most likely explanation for the data is that the reactions follow an induced fit mechanism; for a conformational selection mechanism to still be valid, the mutation would have to result in minimal to no change in the concentration of protein in the correct conformation for binding, which seems unlikely given the widespread changes in residual structure.

When making any point mutation, the amino acid side chain is altered. If the interactions with a partner protein are disrupted, the mutation can have significant effects on binding affinity. A 25 ± 4-fold reduction in affinity was observed for MLL upon mutating P9 to alanine. A simple explanation for this shift in K_d, therefore, is that P9 made interactions with CBP-KIX that were lost upon alanine mutation.

However, proline residues are known to influence the conformation of the preceding residue (MacArthur and Thornton, 1991), and analysis of the bound structure indicated that L8 made interactions with CBP-KIX. Removal of the leucine side chain by mutation to alanine resulted in a 53 ± 8-fold reduction in affinity, which was similar to that observed with the P9A mutant. This was again predominantly due to an increase in k_{off}, suggesting that the decrease in affinity of P9A was due to disruption of the L8:CBP-KIX interaction. A recent mutational analysis of the MLL:KIX interaction found that mutation of the same leucine did not alter the binding affinity at pH 4 (Toto and Gianni, 2016). Although, the authors could not measure the effects at pH 7.2 and noted that several residues showed different behaviours when the pH was altered. The buffer used in this study was at pH 7.4, which could explain the differing result.

Table 4.1 Helicity, Kinetics and Thermodynamics for p53, MLL and c-MYB peptides.

Peptide	Helicity (NMR) (%)	Helicity (CD) (%)	k_{on} ($\mu M^{-1}s^{-1}$)	k_{off} (s^{-1})	K_d (k_{off}/k_{on}) (nM)	K_d (Equilibrium) (nM)
p53 WT	3 (12)	n.d.	25 ± 1	4.40 ± 0.02	176 ± 8	240 ± 30
p53 P27A	6 (30)	n.d.	30.7 ± 0.6	0.37 ± 0.01 0.072 ± 0.005[1]	n.d.	25 ± 3
MLL WT	2 (3.5)	13	24 ± 1	12.4 ± 0.3	530 ± 30	660 ± 60
MLL P21A	n.d.	15	22.0 ± 0.6	12.22 ± 0.08	560 ± 10	700 ± 100
MLL P9/21A	3.5 (4.6)	17	20 ± 14	196 ± 4	9000 ± 6000	17000 ± 2000
MLL L8A	n.d.	12	n.d.	173 ± 6	n.d.	35000 ± 4000
c-MYB WT	22 (51)	25	27.7 ± 0.5	6.38 ± 0.03	230 ± 4	168 ± 9
c-MYB P17A	n.d.	32	28.1 ± 0.9	8.8 ± 0.1	310 ± 10	320 ± 60
c-MYB P17/44A	29 (60)	35	28.8 ± 0.8	8.97 ± 0.08	311 ± 9	270 ± 20

Data for p53 are from Borcherds *et al.* (2014) and Crabtree *et al.* (2017). Data were collected at 5°C for MLL and 10°C for c-MYB (25°C for p53). Due to technical limitations, NMR data for MLL was collected at 10°C. Helicity within the region that becomes helical upon binding are shown in parenthesis. Errors in k_{on} and k_{off} represent the error of the fit, except for MLL P9/21A and L8A, where the error represents the standard error of the mean (SEM). Equilibrium constant errors represent the SEM (n =3). Errors for the K_d calculated from kinetics were propagated using standard methods. [1]Biphasic dissociation kinetics are observed for p53 P27A – both dissociation rate constants are shown. n.d. not determined.

4.7 Summary

Helix-flanking prolines in three IDPs were found to restrict the level of residual structure, but had differing effects on the affinity of the IDP for its partner. For p53, the prolines reduce binding affinity, whereas they have little effect for c-MYB and enhance the affinity of MLL. Yet, amongst the differing changes in K_d, there was a commonality – the changes in binding affinity were predominantly due to a change in k_{off}, not k_{on}. Thus, rather than producing their affect through controlling residual structure and association rates, helix-flanking prolines were found to influence binding affinity through modulating the lifetime of the bound complex.

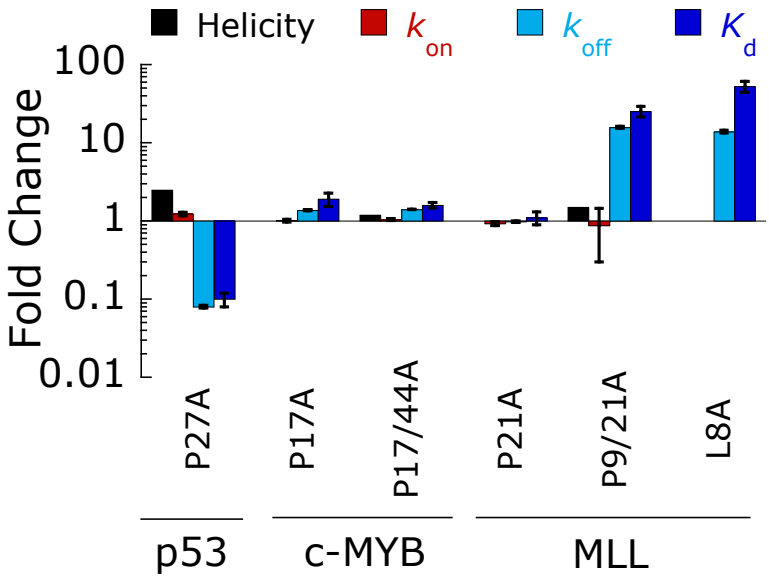


Figure 4.17 Fold Changes in Helicity, Rate Constants and Affinities for p53, MLL and c-MYB Peptides upon Helix-Flanking Proline to Alanine Mutation. Changes in helicity were determined from NMR and represent the fold change within the region that becomes helical upon binding. MLL P21A, L8A and c-MYB P17A were not investigated by NMR; therefore, helicity data is not shown for these peptides. Errors were propagated using standard methods.

Chapter 5

Modulation of Residual Structure in IDPs

The Genetic Algorithm for Design of Intrinsic Secondary Structure (GADIS) was devised and implemented by Tyler Harmon, a PhD student in Professor Rohit Pappu's lab (University of Washington, St. Louis). Peptide sequences were designed by Tyler, using GADIS. Kinetic data were collected and analysed together with Dr Sarah Shammas.

5.1 Methods to Alter Residual Structure

One way to address the role of residual structure in IDP coupled folding and binding reactions is to alter the level of IDP residual structure, and follow the impact on the interaction kinetics and thermodynamics (as described in the previous chapter). Various methods are available to modify levels of secondary structure. For example, for IDPs that form helices upon binding, mutation of residues - which are solvent exposed in the bound structure - to proline can reduce residual helical structure (Rogers *et al*., 2014b).

Enhancing helicity is a little trickier. Chemical stabilisers such as trifluorethanol (TFE) or protein modifying strategies such as constraining or stapling can enhance IDP helicity (Lau *et al.*, 2015). Stapling techniques involve substituting 2 positions on the same side of a helix (e.g. *i* and *i*+4) with non-natural amino acids. The side chains of the substituting amino acids contain groups that can interact with each other, forming a covalent linkage between the 2 positions. By optimising the length and position of the linking staple, the sequence between the substituted groups can be constrained into a helix. Cooperativity then promotes helix formation in the surrounding residues.

While useful, all of these methods have issues. TFE can alter the structure of both the IDP and the partner protein. Whereas, proline scanning or constraining the IDP by stapling both require mutations to the amino acid sequence, altering the composition. This is particularly problematic when the mutation results in a change

in net charge: long-range electrostatic interactions can influence association rates, confounding any changes that are due to the alteration of residual structure. Isolating only the effects of altering residual structure therefore required a new method. One that could specifically modulate the helicity of the IDP, without altering the amino acid composition.

5.2 GADIS

Both positively and negatively charges amino acids are overrepresented in IDPs (Romero *et al.*, 2001; Theillet *et al.*, 2014; Uversky *et al.*, 2000). Without altering the total number of charges, or the sequence composition, these charges can be utilised to modulate IDP structure (Das and Pappu, 2013). Arranging the charges in a manner where the positive and negative charges are well mixed produces a random-coil like structure. Whereas, shuffling the sequence into blocks of oppositely charged regions allows long-range electrostatic attraction, creating hairpin-like conformations.

When placed in the middle of an alanine host peptide (AAXAA) sequence, each amino acid (position X) samples similar, but distinct, phi and psi angles (Towse *et al.*, 2016). If the presence of a specific side chain can alter the phi and psi angle preferences (Serrano, 1995), then it follows that each amino acid can differentially influence structure. An additional layer of complication occurs in proteins, where the sequence is more intricate than AAXAA. Interactions between adjacent side chains can result in steric clashes and restrictions on side chain rotation, resulting in a loss of entropy (Creamer and Rose, 1992; Horovitz *et al.*, 1992; Srinivasan and Rose, 1999). Torsion angles that reduce side chain restriction and clashing are favoured; however, favourable inter-side chain enthalpic contributions can compensate for entropic loses. Taken together, this means that the presence and conformation of one amino acid can influence the conformation of neighbouring residues (Keskin *et al.*, 2004; Pappu *et al.*, 2000; Zaman *et al.*, 2003).

The Genetic Algorithm for Design of Intrinsic Secondary Structure (GADIS) takes advantage of both long-range side chain interactions, and the context dependent nature of phi and psi angles. Through changing neighbouring amino acids, the algorithm aims to produce protein sequences with specific helicity profiles (Harmon

et al., 2016). Unlike mutagenesis, the overall sequence composition is maintained, with only the order of amino acids altered.

5.2.1 Altering Residual Helicity in PUMA

The intrinsically disordered BH3–only protein, PUMA folds upon binding to the structured BCL-2-like protein, MCL-1. Proline scanning mutagenesis had been previously employed to investigate the role of residual structure in this coupled folding and binding reaction (Rogers *et al*., 2014b). PUMA was therefore an ideal protein to study using GADIS, as the effects of altering residual helicity could be compared to the previous study.

Altering the position of PUMA residues that form the MCL-1 interaction interface could significantly affect the reaction kinetics and complex affinity. To avoid disrupting key PUMA:MCL-1 interactions, 7 interface residues (W133, I137, L141, I144, A145, L148 and Y152) were kept in their relative WT positions. The remaining 27 residues were shuffled to create 5 different helicity profiles (Figure 5.1).

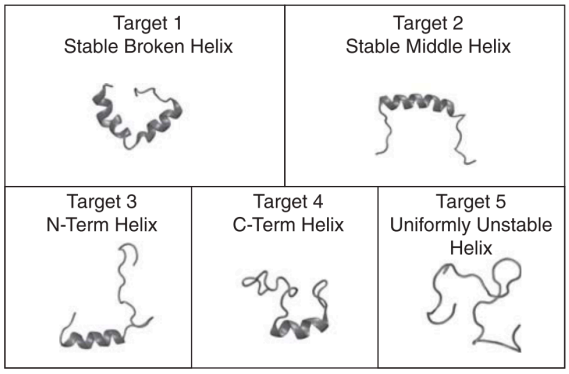


Figure 5.1 Helicity Profiles Targeted Using GADIS.
To investigate the role of residual helical structure in different regions of the protein, PUMA sequences were shuffled in an attempt to produce 5 different target helicity profiles. From this point on, the names of the target helicity profiles are referred to as SBH (stable broken helix), SMH (stable middle helix), NTH (N-term helix), CTH (C-term helix) and UUH (uniformly unstable helix). Figure taken from Harmon *et al.* (2016).

Helicity profiles of the shuffled sequences were determined from all atom ABSINTH-based simulations (Vitalis and Pappu, 2010) and compared to the target helicity

profile. For each profile, 2 sequences were chosen for experimental analysis (Figure 5.2).

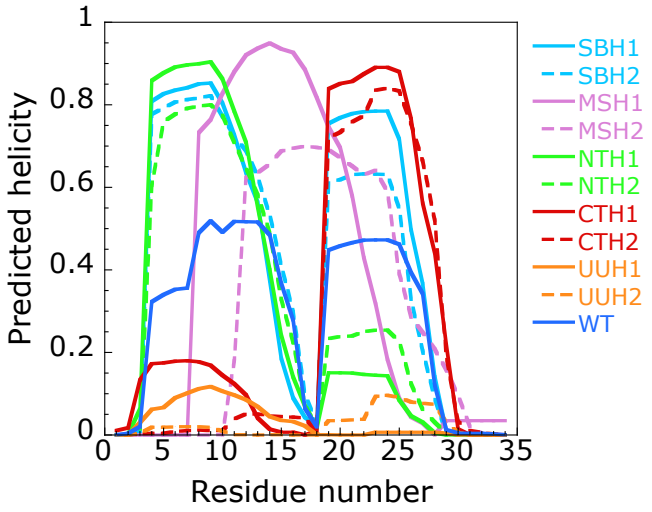


Figure 5.2 Predicted Helicity Profiles for WT PUMA and PUMA Shuffles. For each target profile, 2 sequences were produced by GADIS. The helicity profiles shown above were predicted from all atom ABSINTH-based simulations (Vitalis and Pappu, 2010). WT PUMA resembles a broken helix, with partial helicity between residues 3 – 17, and 18 – 29. The other helicity profiles were targeted to reduce all helicity (UUH); create a single helix spanning the middle of the sequence (MSH); increase helicity between residues 3 – 17 while reducing it between 18 – 29 (NTH); increase helicity between residues 18 – 29 while reducing it between 3 – 7 (CTH); and maintain a WT profile, but with enhanced helicity (SBH).

5.3 Results

Many IDPs that undergo coupled folding and binding can sample the 'bound-like' structure when unbound. A convenient technique to assess the importance of this 'bound-like' structure, is to alter the residual helical content of the IDP, and follow the effect on the binding kinetics and affinity. However, methods to alter residual structure content are typically either non-specific, or require altering the sequence composition through mutagenesis. To further investigate the role of residual helical structure in the IDP, PUMA, a new method based on shuffling amino acids was developed. The ability of this method to alter residual helical content was assessed by ultraviolet-circular dichroism spectroscopy (CD).

5.3.1 GADIS Produces PUMA Sequences with Varying Levels of Helicity

As previously noted (Rogers *et al.*, 2013), WT PUMA 34 formed oligomers within the μM range. The oligomers were observed in CD experiments, with scans performed above 3.5 μM displaying concentration dependent increases in α-helical content (Figure 5.3).

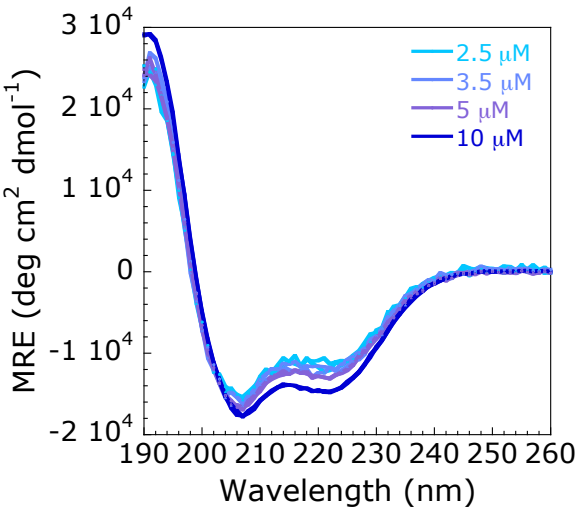


Figure 5.3 Concentration Dependent CD Signal Indicates PUMA 34 Oligomerisation.
The CD signal for WT PUMA 34 is concentration dependent, indicating the formation of homo-oligomers. Below 3.5 μM, the signal is no longer concentration dependent, and the oligomers cannot be detected in CD experiments. For each concentration, 2 independent samples were scanned. Both scans are shown. MRE indicates the mean residual ellipticity.

To determine the success of GADIS, accurate helicity estimates were required. As peptide oligomerisation increased the α-helical content of WT PUMA 34, it was important to ensure that CD scans were representative of the monomeric peptide. Consequently, at least 3 different concentrations were scanned for each shuffled sequence (Figure 5.4). The highest concentration scan that did not show concentration dependence was then used for further analysis.

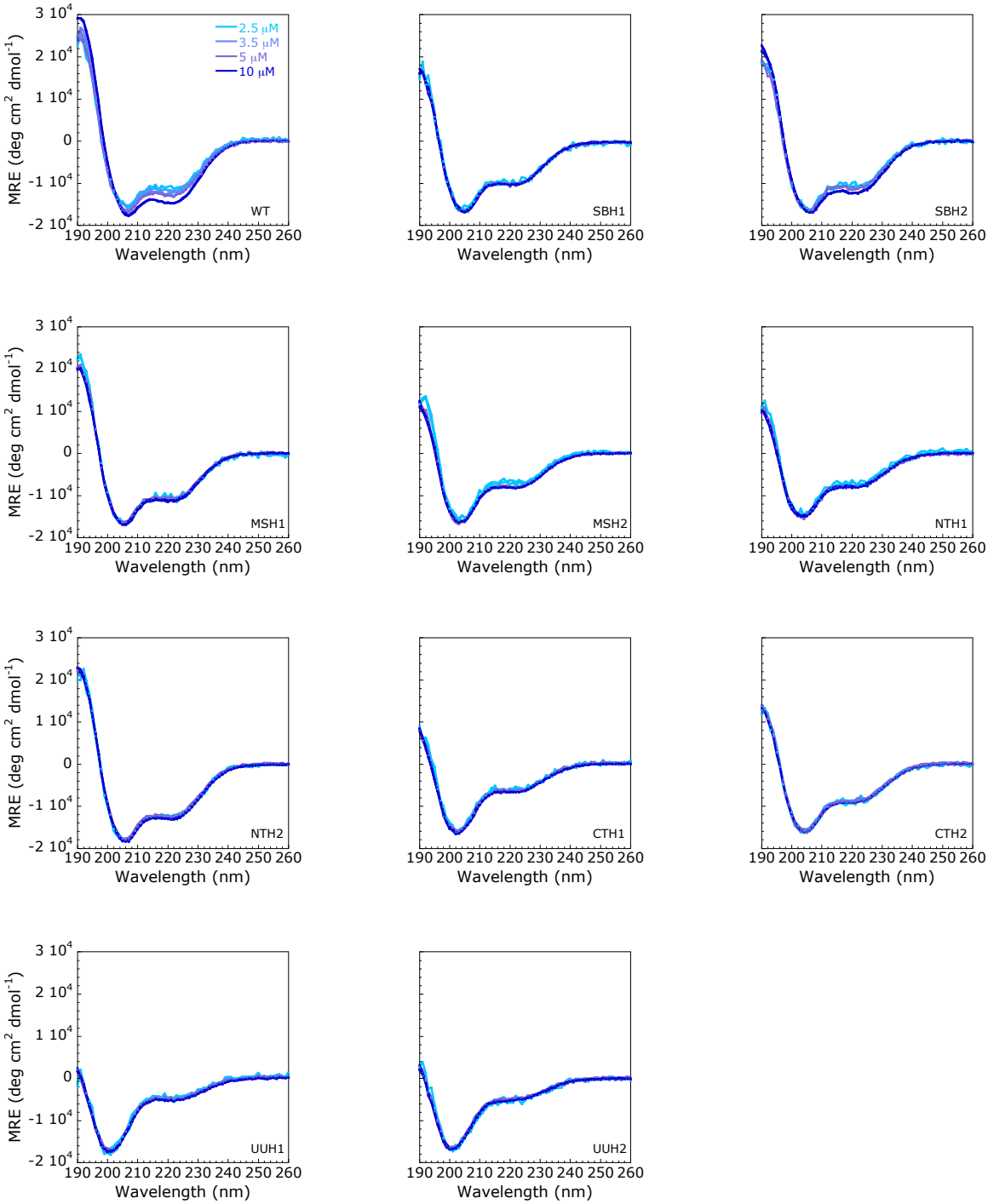


Figure 5.4 CD Concentration Dependence for PUMA 34 Shuffles.
Concentration dependent CD scans were only observed for WT (> 3.5 μM) and SBH2 (> 5 μM). MRE indicates the mean residual ellipticity.

Comparison of the CD spectra indicated that various levels of structure could be produced by shuffling the position of 27 out of the 34 PUMA residues (Figure 5.5).

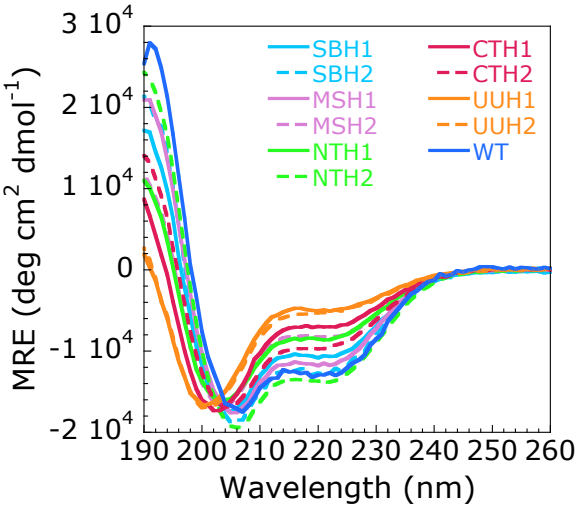


Figure 5.5 Comparison of the CD Spectra for PUMA shuffles.
UUH1 and UUH2 display largely disordered CD spectra (characterised by a trough at approximately 200 nm). Increases in helicity are indicated by a reduction in the mean residual ellipticity (MRE) at 222 nm and a shift in the 200-nm trough towards 208 nm.

To confirm changes in residual structure, helicity estimates were produced from the MRE value at 222 nm. Only NTH2 (39%) showed an increase in total helical content compared to WT PUMA (36%). The helicity of the remaining shuffles varied from 13% (UUH1) to 36% (SHB2). Differences in total helical content were also observed within the same target profile, e.g. NTH1 (24%) *vs* NTH2 (39%).

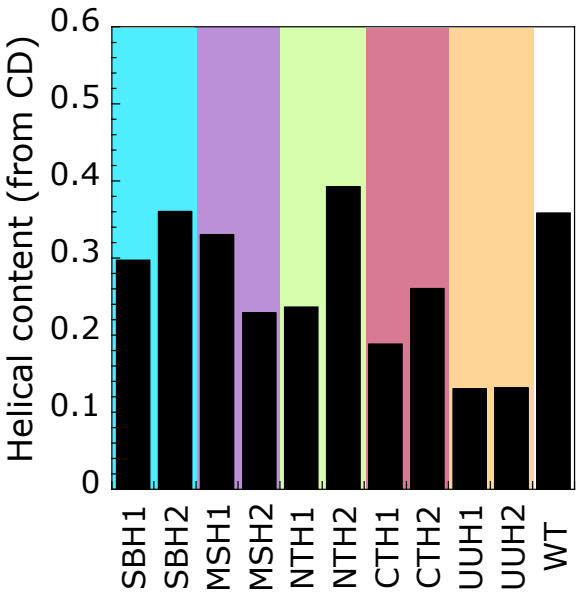


Figure 5.6 Fraction of Helical Content in PUMA Shuffles.
The fraction of total helical content was calculated for each PUMA shuffle using the MRE value at 222 nm and the method of Muñoz and Serrano (1995).

Shuffling the order of amino acids produced proteins with various levels of helicity. However, for GADIS to be successful, the changes in helicity need to be predictable. While the trends are similar, the absolute estimates of structural content from CD are highly dependent on the method used to convert MRE to helicity. Therefore, to allow comparison with the helicity obtained from simulations (Figure 5.7), multiple methods were used to compute helicity (Chen and Yang, 1971; Chen *et al.*, 1974; Greenfield and Fasman, 1969; Morrisett *et al.*, 1973; Muñoz and Serrano, 1995). Each method provided a different absolute value for helical content; however, the Pearson correlation coefficient, obtained from a straight-line fit of the CD vs simulation data, was similar in all cases (0.73 – 0.76).

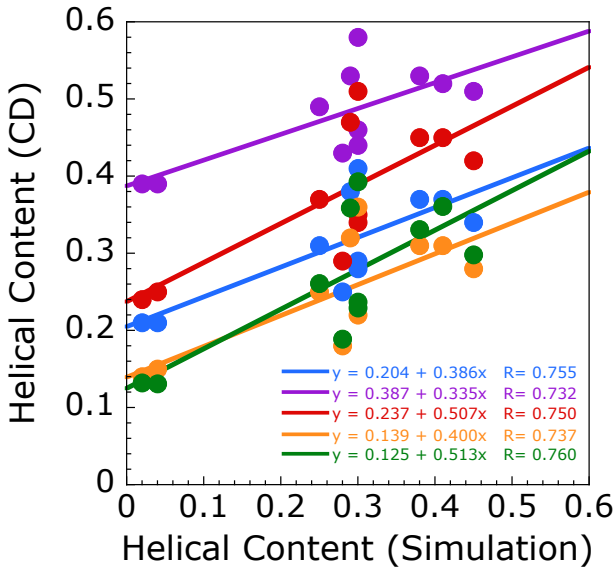


Figure 5.7 Correlation of Fractional Helicity Calculated from CD and Simulations.

Helicity from CD was calculated using the methods of Greenfield *et al.* (1969) (purple); Chen *et al.* (1971) (red); Morriset *et al.* (1973) (blue); Chen *et al.* (1974) (yellow); and Muñoz and Serrano (1995) (green). The data are fit to a straight line, with the equation and the Pearson correlation coefficient (R) shown.

5.3.2 PUMA Shuffles Alter the Interaction with MCL-1

GADIS successfully produced shuffled sequences with varying levels of helicity. Upon binding to MCL-1, WT PUMA folds to form a single contiguous α-helix (Figure 5.8). By altering the structure of the free peptide, the folding upon binding reaction could have been affected, leading to a difference in the structure of the bound complex. To check that the altered conformation of the free peptide had not disrupted the structure of the PUMA:MCL-1 complex, CD was utilised (Figure 5.9). Only CTH1 displayed a spectrum that resembled WT PUMA:MCL-1. All other PUMA shuffles either had no change in structure compared to the 2 free proteins (UUH1 and SBH1), or had an intermediary spectrum, which occurred between WT PUMA:MCL-1 and the spectrum that would be expected if there was no change in structure upon binding.

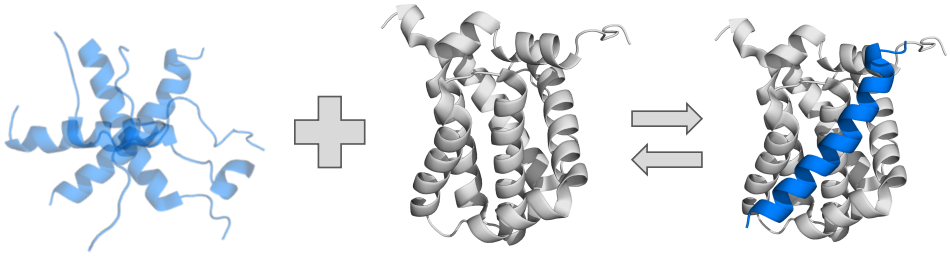


Figure 5.8 The Disordered Protein PUMA Undergoes Coupled Folding and Binding with its Partner, MCL-1.
Unbound PUMA (blue) is disordered an exists an ensemble of conformations. Upon binding, PUMA forms a single contiguous α-helix. Structures were generated using generated using PyMOL (version 1.7.2.1, Schrödinger). MCL-1 (grey) and the PUMA:MCL-1 complex are based on PDB code 2ROC. The PUMA peptide used to determine the structure was 27 amino acids (residues 130 – 155 with an additional C-terminal methionine), compared to the 34-mer peptide investigated in this chapter (residues 128 – 161). In the PUMA:MCL-1 complex, the N-terminus of PUMA is at the bottom left of the structure.

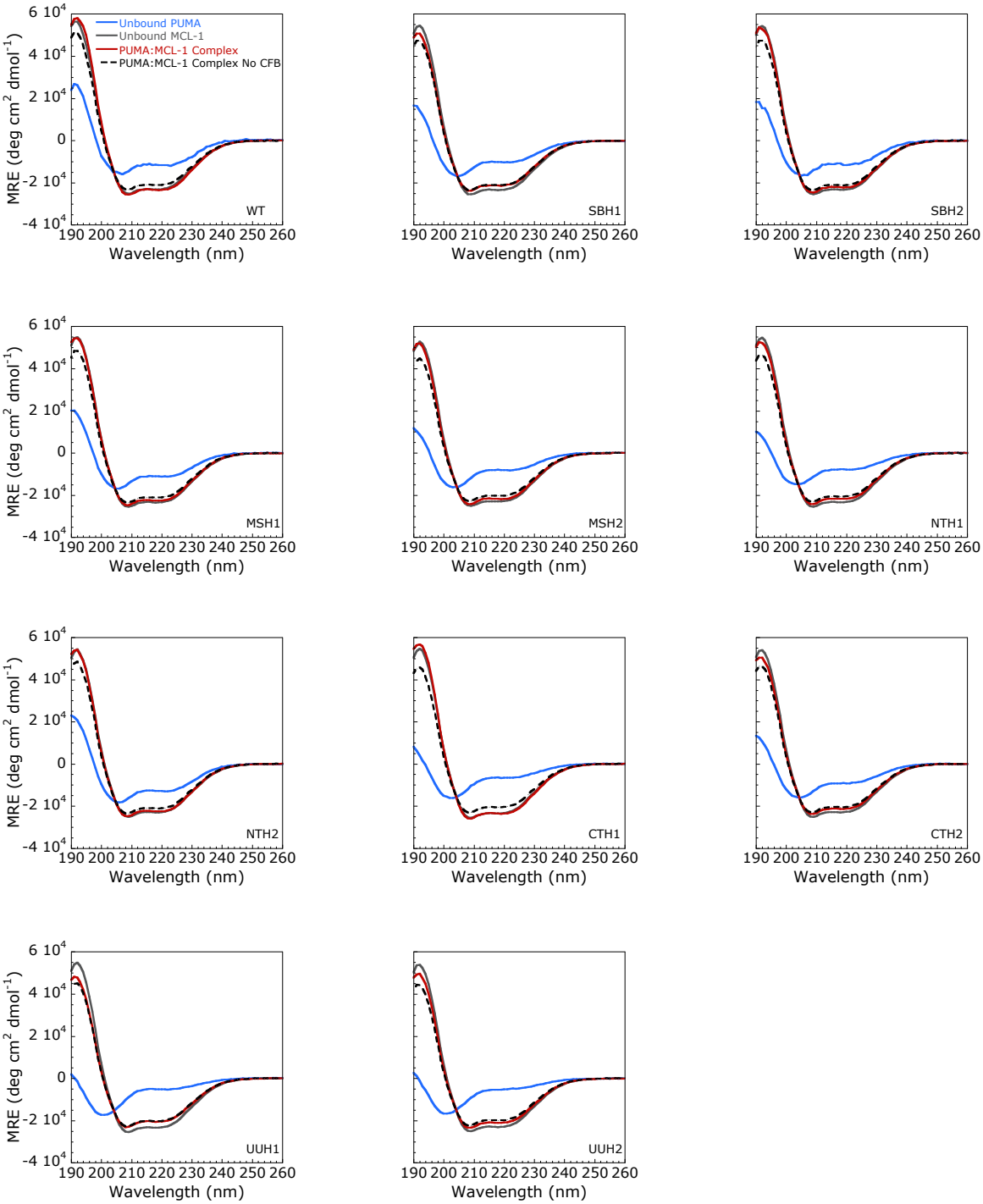


Figure 5.9 Coupled Folding and Binding of PUMA Shuffles Probed by CD. Protein solutions used to obtain the PUMA (shuffle) and MCL-1 spectra were mixed in a 1:1 volume and concentration ratio. Final concentrations were between 1.75 and 2.5 μM. The expected spectrum (black, dashed line), if no folding upon binding occurred (no CFB), was calculated from the mean of the individual spectra. Spectra for PUMA:MCL-1 mixtures (red) gained various amounts of helicity compared to the expected spectrum of the free proteins.

The differences in the bound spectra indicated that there was a difference in the bound complex. Either the structure of the bound state was altered. Or not all of the PUMA shuffle had gone into complex with MCL-1. The K_d of WT PUMA:MCL-1 is 0.1 ± 0.03 nM (Rogers *et al.*, 2013). As both PUMA and MCL-1 were at a concentration in the low micro molar range for the CD experiments, to be consistent with an explanation relating to unsaturated binding, the K_d for the shuffled peptides would have to be shifted by approximately 4 orders of magnitude. This would bring the K_d into a range that is easily measureable using a reversible bimolecular kinetic experiment (Equation 2.26), where both the k_{on} and K_d can be extracted. For WT PUMA, a similar amplitude change was observed when the lowest (0.2 μM) and highest concentrations (1 μM) were mixed with 0.2 μM of MCL-1 (Figure 5.10). Only the timescale of the reaction was altered. On the other hand, for all of the measured PUMA shuffles, an increase in amplitude was observed when the concentration of PUMA was increased (e.g. UUH2 in Figure 5.10). Note that no change in intrinsic tryptophan fluorescence was observed for some PUMA shuffles (SBH1, CTH2 and UUH1). Consequently, the k_{on} and K_d for these three proteins could not be determined using the bimolecular reversible association method.

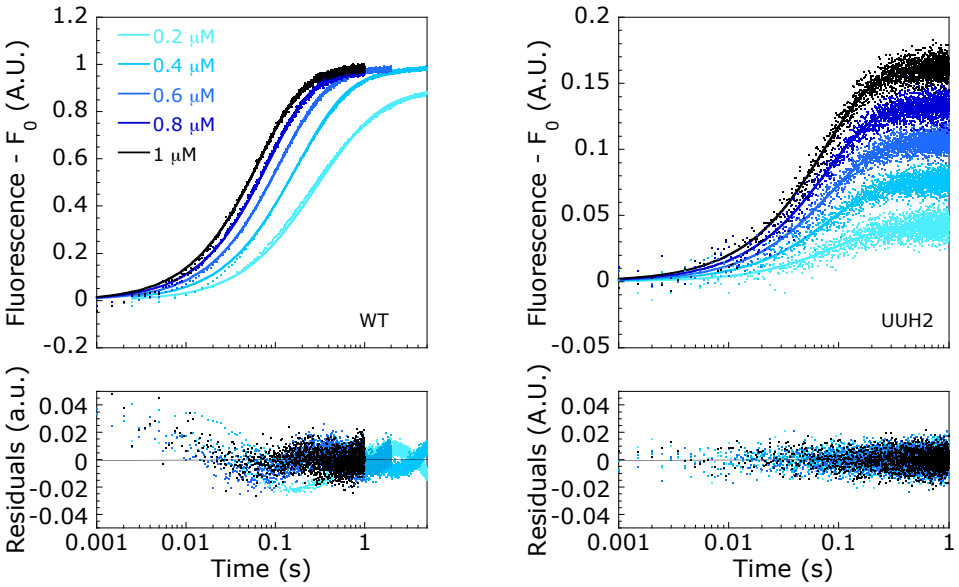


Figure 5.10 Examples of Kinetic Traces for PUMA Shuffles.
Various concentrations of PUMA (0.2 – 1 μM) were rapidly mixed in a 1:1 volume ratio with 0.2 μM of MCL-1 using stopped-flow. The change in intrinsic tryptophan fluorescence with time was followed and the data were globally fit to a reversible bimolecular association model (Equation 2.26). For some proteins, a linear term was added to the model, to account for photo-bleaching and/or the presence of PUMA oligomers (as described previously in Rogers *et al.* 2013). To aid comparison of the changes in amplitude, the fluorescence at time 0 (obtained from the fit) was subtracted from the observed fluorescence. Data for UUH2 demonstrate the increase in amplitude observed for the PUMA shuffles. While the k_{on} can be determined, the K_d for WT PUMA was shown previously to be 0.1 ± 0.03 nM (Rogers *et al.*, 2013), which is too tight to accurately measure using this technique.

As the experiments were performed with a minimum of 0.2 nM of protein, MCL-1 would be expected to go fully into complex with any shuffles that had a similar K_d to WT PUMA. The fluorescence amplitude change would therefore be similar for all concentrations of PUMA, as observed for WT. However, when mixed with higher concentrations of each PUMA shuffle, the fluorescence amplitude change increased, indicating that different amounts MCL-1 were going into complex. Analysis of the data indicated that this was due to a reduction in binding affinity (Figure 5.11). As suggested by the CD data (Figure 5.9), CTH1 had the tightest binding affinity of all the shuffles, although the K_d was still shifted by > 2 orders of magnitude compared to WT. At the other end of the spectrum, NTH2 had the lowest affinity for MCL-1, with a shift in K_d of > 4 orders of magnitude.

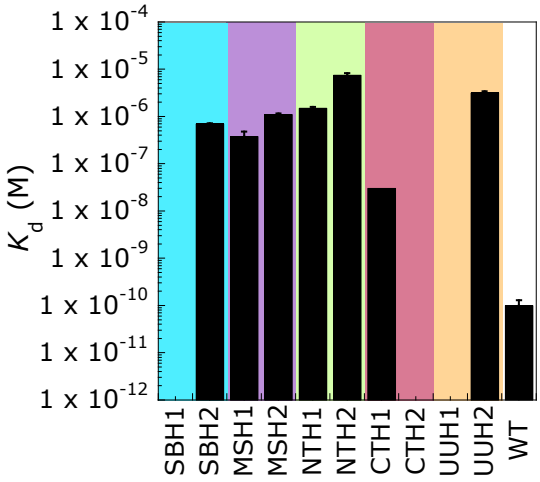


Figure 5.11 PUMA Shuffles Have Reduced Affinity for MCL-1.
The K_d was determined from reversible bimolecular association experiments. Error bars represent the error of the fit, except for WT PUMA, where the error bars represent the standard error of the mean. Due to a lack of signal change, data were not collected for SBH1, CTH2 or UUH1. Data for WT was obtained from Rogers *et al.* (2013).

The experimental set up allows the cause of the shift in K_d to be examined. Apart from NTH2, the association rate constant for all the shuffles and WT were approximately within the same order of magnitude (10^6 – 10^7 $M^{-1}s^{-1}$) (Figure 5.12). Clearly, the 2 to 4 orders of magnitude shift in K_d could not be explained by a change in k_{on}. Assuming a 2-state system, the k_{off} could be estimated by multiplying the K_d by the k_{on}. WT PUMA has a dissociation rate constant of 0.0016 ± 0.0005 s^{-1} (Rogers *et al.*, 2013). In contrast the k_{off} of the slowest dissociating mutant was 0.360 ± 0.001 s^{-1} (CTH1), while at the other end of the spectrum, the k_{off} for MSH2 (30 ± 2 s^{-1}) was shifted by > 4 orders of magnitude. Calculating the k_{off} in this way has an issue – it is not a direct measurement, so errors can be compounded. To confirm the changes in k_{off}, dissociation rate constants were obtained from out-competition experiments for 3 of the peptides. These matched the calculated data reasonably well.

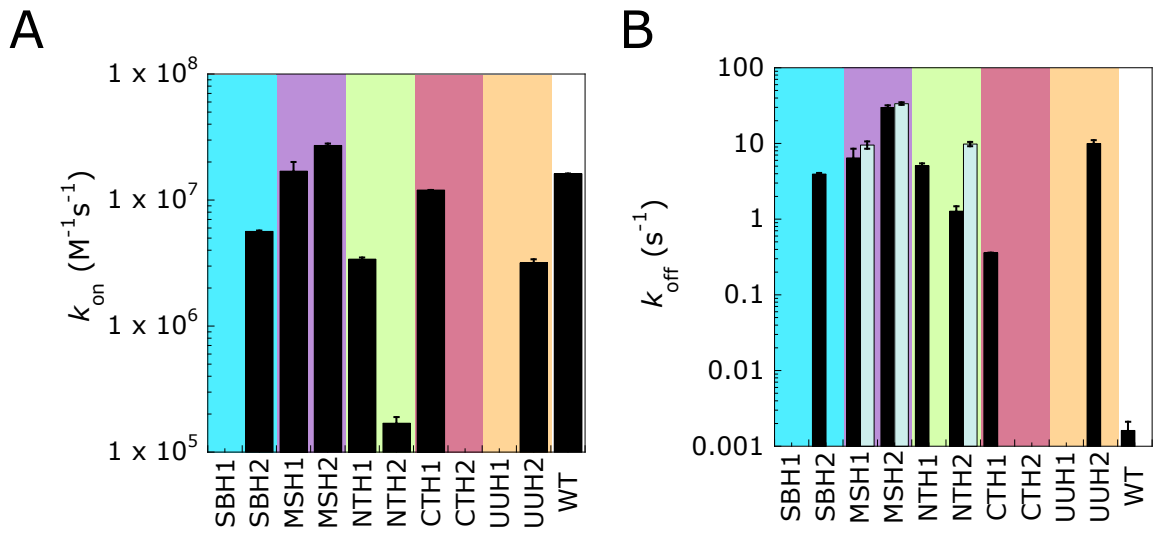


Figure 5.12 Shift in K_d Predominantly Due to a Change in k_{off}.
(A) Reversible bimolecular association experiments demonstrated that all shuffles, except NTH2, bound at a similar rate. The difference in K_d was instead due to significant changes in k_{off} (B). k_{off} was determined by multiplying K_d by k_{on} (black bars). For 3 shuffles, the calculated k_{off} was compared to the k_{off} determined from out competition experiments (grey bars). Due to a lack of signal change, data were not collected for SBH1, CTH2 or UUH1. Error bars represent the error of the fit, or the standard error of the mean (out competition k_{off}). Data for WT k_{off} was obtained from Rogers *et al.* (2013).

5.3.3 Investigating the Cause of the Shifts in Affinity and k_{off}

A change in k_{off} indicated that the shuffles were mainly affecting the bound state of PUMA:MCL-1, rather than the rate of complex formation. Disruption of the bound complex could occur through disturbing PUMA WT-like interactions with MCL-1. Although 7 hydrophobic residues were kept in the same position as WT, it was possible that some important contacts were lost during the shuffling process, or that unfavourable interactions were introduced. Fortunately, the structure of PUMA bound to MCL-1 has been solved by NMR (Day *et al.*, 2008). Based on the PDB file, contact map analysis allows interacting residues in different protein chains to be identified (Sobolev *et al.*, 2005). Utilising this technique, 19 PUMA residues were identified to interact with MCL-1 (Figure 5.13). As this was considerably more than the 7 that were kept constant in the PUMA shuffles designs, it was possible that the altered affinity was due to a loss of some of these 19 contacts. New peptide designs, with all 19 contacting residues maintained and the other 15 shuffled, were produced to test this hypothesis.

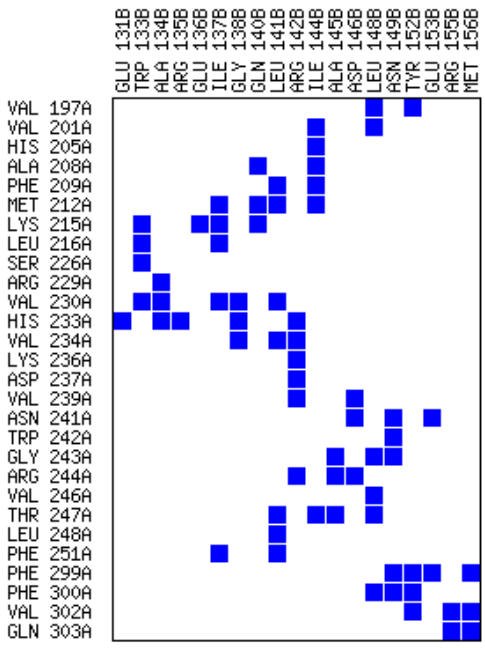


Figure 5.13 Contact Map Analysis of PUMA Bound to MCL-1.
A contact map analysis was produced using the PDB structure, 2ROC and the SPACE suite web tool (Sobolev *et al.*, 2005). Contacts were defined as having a contact area of > 11 Å^2, which was the default setting. In total, 19 contacts were found between PUMA (chain B, top) and MCL-1 (chain A, side). Out of these 19, 7 were kept in the same relative position in each of the shuffle designs (TRP 133, ILE 137, LEU 141, ILE 144, ALA 145, LEU 148 and TYR 152).

The conserved shuffles displayed similar levels of residual helicity, but were all less helical than WT PUMA (conserved 1 – 23%; conserved 2 – 26%; conserved 3 – 22%; and WT – 36% estimated helicity). Conserved shuffle 3 also displayed concentration dependent CD spectra above 5 μM, indicating the presence of oligomers. Mixtures of the conserved shuffle peptides with MCL-1 all displayed an increase in total α-helical content (Figure 5.14).

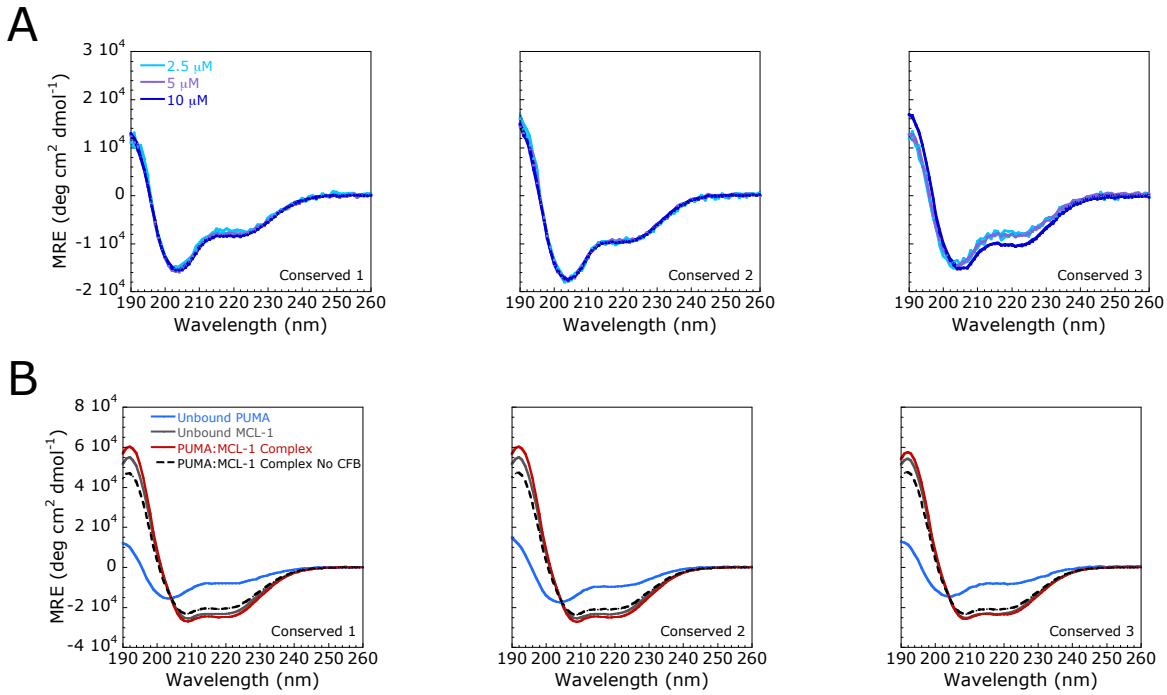


Figure 5.14 CD Spectra for Conserved PUMA Shuffles.
(A) Concentration dependence for the CD spectra of unbound shuffles. Each shuffle has 19 residues that are 'conserved' in the same relative position as WT PUMA. Conserved shuffle 3 shows a concentration dependent CD spectrum above 5 μM, indicating the presence of oligomers. (B) A gain in total α-helical content was observed for all 3 conserved shuffles upon binding to MCL-1. The expected spectrum (black, dashed line), if no folding upon binding occurred (no CFB), was calculated from the mean of the individual spectra.

The affinity of the conserved shuffles for MCL-1 was much closer to WT PUMA than to the previous shuffle designs (Figure 5.15). However, the K_d for conserved shuffle 3 (2.94 ± 0.04 nM) was still shifted by more than an order of magnitude compared to WT PUMA (0.1 ± 0.03 nM). The affinity of the other conservative shuffles was also reduced by approximately 7-fold compared to WT (conserved 1 K_d = 0.73 ± 0.02 nM; conserved 2 K_d = 0.75 ± 0.02 nM). Reductions in affinity were not due to changes in k_{on}, with all conserved shuffles displaying association rate constants in the 10 – 20 $\mu M^{-1}s^{-1}$ range (WT k_{on} = 16.19 ± 0.02 $\mu M^{-1}s^{-1}$).

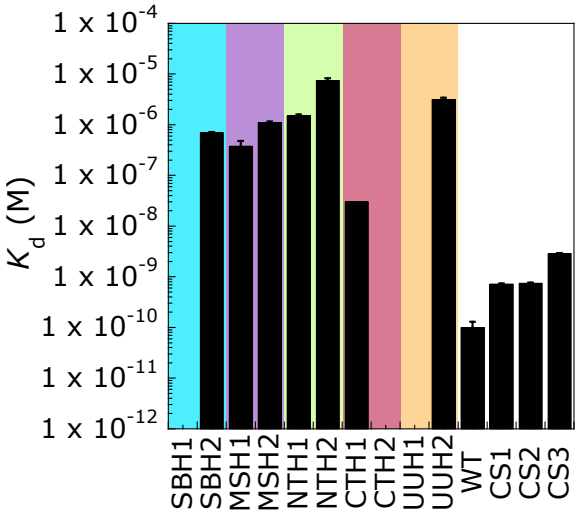


Figure 5.15 Affinity of Conserved Shuffles for MCL-1 Compared to WT and the Previous Shuffle Designs.
The K_d was determined from reversible bimolecular association experiments. Error bars represent the error of the fit, except for WT PUMA, where the error bars represent the standard error of the mean. Due to a lack of signal change, data were not collected for SBH1, CTH2 or UUH1. Data for WT was obtained from Rogers *et al.* (2013). The experiment also provided association rate constants of 17.68 ± 0.01 $\mu M^{-1}s^{-1}$ for conserved shuffle 1 (CS1), 15.58 ± 0.01 $\mu M^{-1}s^{-1}$ for conserved shuffle 2 (CS2) and 15.16 ± 0.01 $\mu M^{-1}s^{-1}$ for conserved shuffle 3 (CS3).

5.3.3.1 Requirement for a Small Residue in PUMA at Position 138

Keeping all 19 contacting residues in a fixed position had brought the affinity of the PUMA shuffles closer to the WT K_d, suggesting that some contacts may have been lost in the original shuffle designs. To investigate whether there were any obvious contacts that could have been lost, sequences of WT BH3 containing proteins were aligned. Only proteins that bind to MCL-1 with similar affinity to PUMA were chosen for analysis. Just 3 residues were identical in all 4 proteins and only an additional 9 demonstrated strongly or weakly similar properties (Figure 5.16). Out of the 12 residues, 5 (including all 3 identical residues) were already kept constant in the original shuffles. Of the remaining 7 residues, one stood out. In position 138 in PUMA a glycine residue is present that, upon binding, becomes buried into the hydrophobic grove of MCL-1. In all of the other 3 BH3 domains, a small residue (alanine or glycine) is present in this position. However, in 7/10 of the original shuffle designs, this residue was shuffled to a charged residue (either a glutamic acid or an arginine) (Figure 3.9). Burying an unpaired charged residue away from

the solvent (H_2O) can cause a significant energetic penalty upon binding. To investigate whether shuffling G138 was the reason for the 2 – 4 orders of magnitude reduction in K_d for the original shuffles (Figure 5.11), 2 further peptides were designed. The first was based on CTH1, the tightest binding of the shuffles (Figure 5.9 and Figure 5.11). In CTH1, G138 was shuffled with a glutamic acid. This glutamic acid was swapped back into the position of the glycine, with the glycine returning to the WT PUMA position. A second design was based on WT PUMA: a glutamic acid (E130) from the N-terminal flanking region was swapped with the glycine (Figure 3.10).

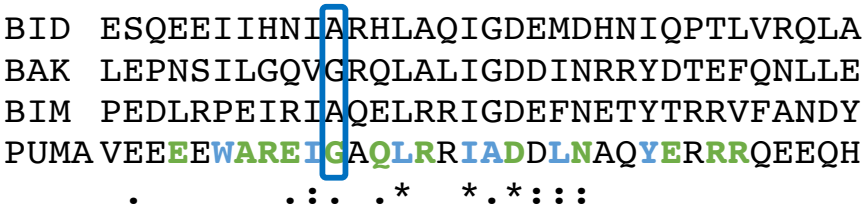


Figure 5.16 Alignment of BH3 Domains from Proteins that Bind with High Affinity to MCL-1.
The BH3 domains from BID, BIM and BAK all bind to MCL-1 with a similar affinity to PUMA (Ku *et al.*, 2011). Sequences were aligned using Clutsal Omega (Goujon *et al.*, 2010; Sievers *et al.*, 2014). As determined by Clustal Omega, positions of full residue conservation are indicated by an asterisk, a colon indicates conservation of strongly similar amino acid properties, weakly similar properties are specified with a period. Residues that were kept constant in the original shuffles are shown in blue. Additional residues that were kept constant in the conserved shuffles are shown in green. The conservation of a small residue at PUMA position 138 is indicated.

Swapping E130 with G138 in the context of WT PUMA increased the helicity from 36% to 54% (Figure 5.17). To test if this due to an oligomeric species, the concentration dependence of the CD signal was examined. No concentration dependence was observed, indicating that either the increased helicity was not due to oligomer formation, or that the oligomeric species was more stable and the population did not change upon altering the concentration in the micro molar range. On the other hand, swapping the glycine back into the WT position in the context of the CTH1 sequence had little impact on helicity (20% *vs* 19% for CTH1 G swap and CTH1, respectively). Although the CD spectra for the original CTH1 sequence

did not display any concentration dependence (Figure 5.4), oligomerisation was observed above 5 μM when the glycine was swapped back to the WT position (Figure 5.17). Upon mixing with MCL-1, a gain in α-helical structure was observed for both WT G swap and CTH1 G swap.

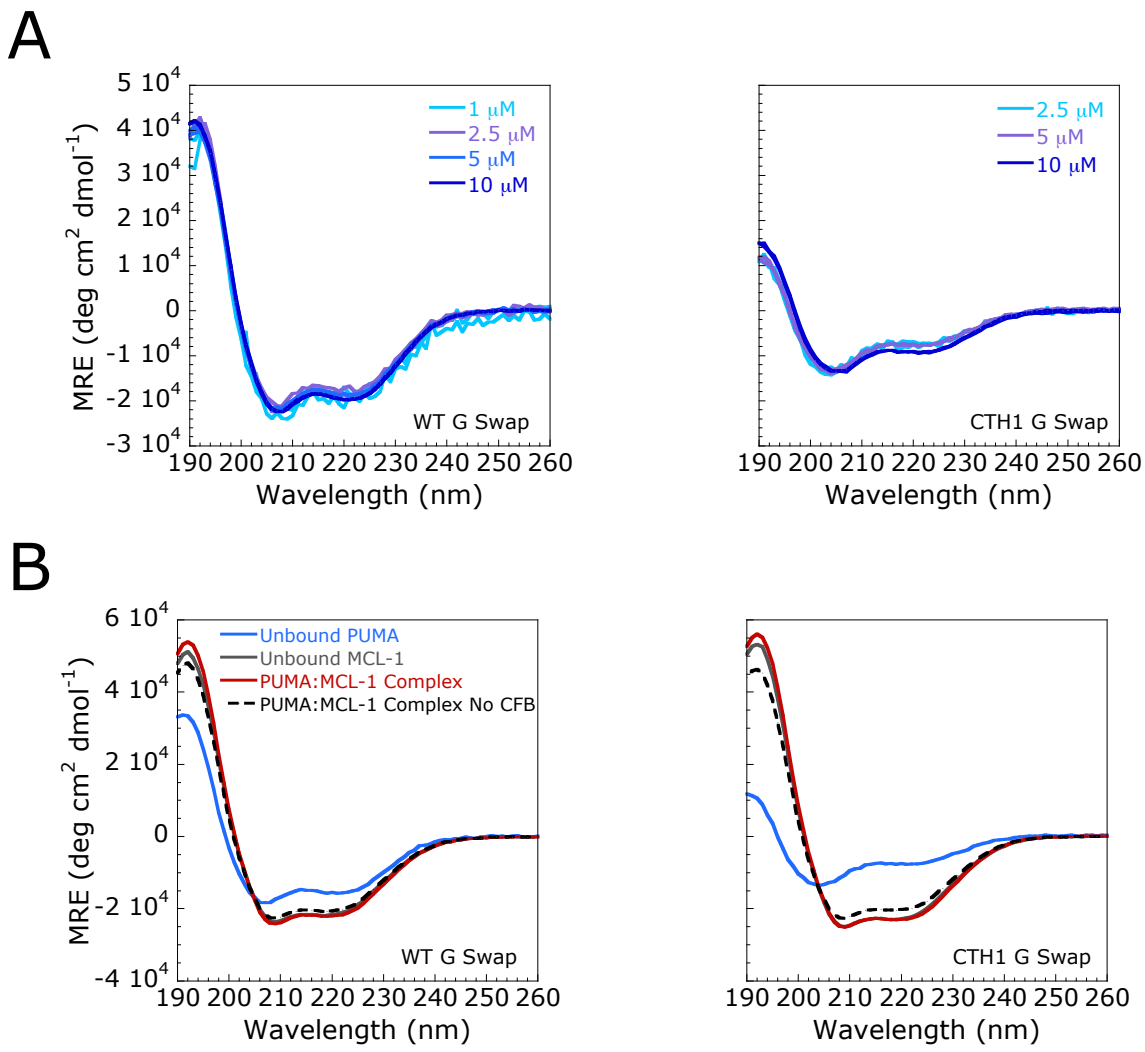


Figure 5.17 CD Spectra for G Swap PUMA Shuffles.
(A) Concentration dependence for the CD spectra of unbound PUMA G swap shuffles. CTH1 G swap shuffle shows a concentration dependent CD spectra above 5 μM. (B) A gain in total α-helical content was observed for both G swap shuffles upon binding to MCL-1. The expected spectrum (black, dashed line), if no folding upon binding occurred (no CFB), was calculated from the mean of the individual spectra.

Swapping G138 with E130 reduced the affinity for MCL-1 by an order of magnitude (Figure 5.18). A similar order of magnitude increase in affinity was observed for

CTH1 when the glutamic acid at position 138 was swapped with the naturally occurring glycine.

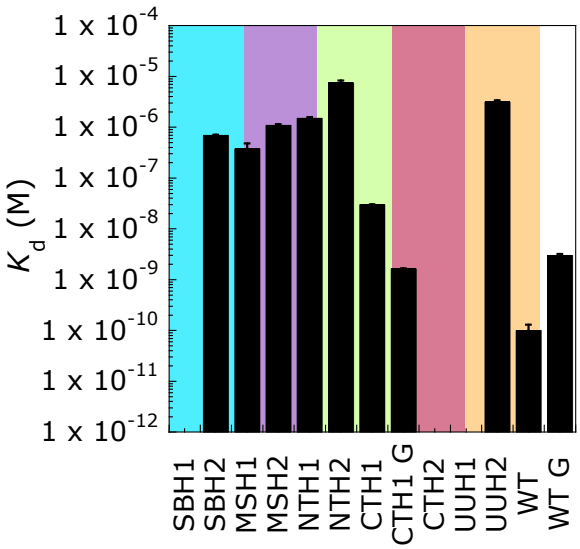


Figure 5.18 The Presence of G138 Modulates the Affinity of WT and CTH1. The affinity of CTH1 G-swap could not be determined using the reversible bimolecular association experiment. Instead, Dr Jeffrey Hollins performed isothermal titration calorimetry (ITC) to ascertain the K_d. The affinity of WT G swap for MCL-1 was investigated using both methods, which showed good agreement (K_d = 3 ± 0.2 nM from ITC and 2.2 ± 0.05 nM from the kinetic experiment). Due to a lack of signal change, data were not collected for SBH1, CTH2 or UUH1. Error bars represent the error of the fit. Data for WT was obtained from Rogers *et al.* (2013).

5.4 Discussion

In line with the view that amino acids can influence the conformation of their neighbouring residues (Keskin *et al.*, 2004; Pappu *et al.*, 2000; Zaman *et al.*, 2003), shuffling of the PUMA sequence produced proteins with varying levels of helicity. GADIS aims to produce PUMA variants with different, but specific helicity profiles. While confirming that the achievement of the target profile requires a residue level technique such as NMR, the determined total helical content from CD correlated reasonably well with the predicted helicity from simulation. GADIS therefore provides a method to alter levels of residual helicity, without altering the amino acid composition. This is a distinct advantage over other techniques, such as proline scanning or peptide stapling, as changes in amino acid composition can alter affinity through mechanisms unrelated to the residual helical content. For

example, mutation of a charged residue to a proline can change the protein's net charge, altering long-range electrostatic interactions and the association rate constant.

In an attempt to prevent interactions between PUMA and MCL-1 being lost during the shuffling procedure, 7 residues were initially kept constant. When the position of the remaining 27 residues was shuffled, the affinity of all of the 10 designed sequences was reduced by between 2 and 4 orders of magnitude. Aligning the sequence of BH3 proteins that bind with high affinity to MCL-1 highlighted the conservation of a small residue at the equivalent position of residue 138 in PUMA. Upon binding, this residue becomes buried in the hydrophobic grove of MCL-1. Typically, during the design process, G138 was swapped with a charged residue (e.g. glutamic acid). Burying an unpaired charged residue into a hydrophobic environment can destabilise a protein, or a protein:partner interaction, by up to ~4 kcal mol^{-1} (Fersht *et al.*, 1985; Tissot *et al.*, 1996). This would match well with the 2 – 4 orders of magnitude shift in K_d observed with the original shuffle designs. However, swapping the G138 with a glutamic acid only produced an approximate 1 order of magnitude (~ 1.3 kcal mol^{-1}) change in affinity. Clearly other factors were involved in the destabilisation of the PUMA shuffle:MCL-1 complex.

Keeping all 19 contacting residues constant indicated that the shift in affinity was predominantly due to a loss of contacts; although, the conserved shuffles did not regain WT affinity. A difference of up to 1 order of magnitude was still observed. This scale of affinity loss is comparable to mutating one of the 7 hydrophobic interacting residues (e.g. W133F or I137A) that were kept constant in the original shuffle designs (Rogers *et al*., 2014a). As reported previously for PUMA mutants binding MCL-1, the majority of the changes in affinity were due to increases in the k_{off}, not k_{on} (Rogers *et al*., 2014a, 2014b). This was still the case for the conserved shuffles. As each one had all 19 contacting residues in the same position, the change in affinity should not arise from a loss of contacts. Additionally, the disparity between the 3 conserved shuffle sequences reduces the probability that a destabilising interaction was introduced in all cases. If neither a loss of stabilising interactions, or the introduction of destabilising interactions can completely explain the shift in affinity for the conserved shuffles, then what is the cause? An intriguing hypothesis can be made – the free energy provided by contacting residues is

dependent on the context of the surrounding amino acid sequence. This is discussed in Chapter 6.

Chapter 6

Influence of Non-Contacting Residues on Binding Affinity

When a protein binds to a partner macromolecule, some of the residues may form contacts that contribute a large proportion of the free energy. Others may still make contacts, but with a lower energy contribution, or not form any interactions with the partner. This raises an interesting question – are the non-contacting residues merely linkers that bring the contacting residues together? Or do they influence the affinity of the interaction? In the previous chapter, shuffling the non-contacting residues in the BH3-only protein, PUMA reduced the affinity for its partner protein, MCL-1 by up to an order of magnitude (Chapter 5.3.3). In those experiments, the sequence composition of PUMA was kept the same, and the relative position of the non-contacting residues was altered.

The BH3-only – BCL-2 like interaction network provides a good system to further investigate the role of non-contacting residues. Multiple BH3-only proteins fold into a single α-helix upon binding to their partner BCL-2 like protein. Interestingly, multiple BH3-only proteins bind with high affinity to MCL-1; however, they share only 9% sequence identity (Figure 5.16). This is far fewer than the 55% of residues that were maintained in identical positions in the PUMA conserved shuffles (Chapter 5.3.3). Unlike the PUMA shuffles, the position and identity of functionally important non-contacting residues will have been subjected to evolutionary constraints. Therefore, multiple options to reach the same goal of high affinity binding could have been sampled. Conservation of the contacting residues need not have occurred, if mutations of the non-contacting residues compensated for any change in binding affinity. In this case, the affinity of the contacting residue would be dependent on the identity of the non-contacting residue(s). This raises an intriguing question – what would happen if the non-contacting residues of one high affinity BH3-only protein were swapped into the same relative position in another BH3-only protein? If the affinity of contacting residues were dependent on the sequence context (i.e. the position and context of non-contacting residues), the hypothesis

would be that swapping the residues would alter the binding affinity. An advantage of this method is that the swapped residues would be present in the same relative positions that they occupy in the original protein. The experiment is similar to the core swap method previously employed to study protein folding mechanisms (Billings *et al.*, 2008; Wensley *et al.*, 2012).

When swapping residues from different proteins, it is possible that some of the amino acids may be charged. Resultant changes in net charge are therefore possible, which could alter the long-range electrostatic interactions, and consequently the k_{on}. To determine how the balance of positively and negatively charged residues would shift upon swapping the non-contacting residues, a number of BH3-only sequences were examined. Aligning with the sequence of PUMA indicated that swapping the residues of BIM and BIK would create a charge balance of -3. Although the charge of PUMA 34 is -4, PUMA 35 contains an additional arginine at the N-terminus (R127), bringing the charge balance to -3. PUMA core swaps were thus designed based on the 19 PUMA residues that contact MCL-1 (Figure 5.13), and the remaining residues from either BIM or BIK (Figure 6.1). Reducing the charge of PUMA to -3 would likely reduce the electrostatic repulsion between PUMA molecules, promoting the homo-oligomerisation observed for PUMA 34 (Figure 5.3). To counteract this effect, an I144A mutation, which is known to reduce the oligomerisation propensity of PUMA (Rogers *et al*., 2014b), was included.

```
BIK:          LMEAVEGRNQVALRLAAIGDEMDLALRSPRLVQLP      5 N, 4 P; overall = -1
BIK Swap:     LMEAEEWAREIGLQLRAAADELNLAYESRRLVQLP      7 N, 4 P; overall = -3
PUMA:         RVEEEEWAREIGAQLRRAADDLNAQYERRRQEEQH      10 N, 7 P; overall = -3
BIM Swap:     EPEDERWAREIGQQLRRAADELNETYERRRFANDY      10 N, 7 P; overall = -3
BIM:          EPEDLRPEIRIAQELRRIGDEFNETYTRRVFANDY      9 N, 6 P; overall = -3
                *      .:. .*   .*:::         :
```

Figure 6.1 Sequences and Charge of BH3-Only Proteins and Core Swaps. Sequences were aligned using Clutsal Omega (Goujon *et al.*, 2010; Sievers *et al.*, 2014). As determined by Clustal Omega, positions of full residue conservation are indicated by an asterisk, a colon indicates conservation of strongly similar amino acid properties, weakly similar properties are specified with a period. The number of negative (N), positive (P) and the difference in charges (overall) are shown. Residues from WT PUMA are shown in blue. Residues from BIM and BIK in the core swaps are indicated in bold.

6.1 Results

The effect of swapping the non-contacting residues of PUMA with residues from BIM and BIK was analysed. The swapped peptides were compared both to WT PUMA, and to WT BIK and BIM.

6.1.1 Core Swaps Alter Residual Secondary Structure

The CD spectra of mouse PUMA 35 (29%), BIM (11%) and BIK (11%) indicated various levels of residual helicity (Figure 6.2). Keeping the 19 MCL-1 contacting residues of PUMA and swapping the equivalent, non-contacting residues of BIM resulted in a peptide with intermediate helicity (18%). Surprisingly, when the core residues of PUMA were swapped into the BIK sequence, the CD spectrum resembles that of an anti-parallel β-sheet. The BIK-swap spectra were concentration independent across a range of 2.5 – 10 μM, suggesting that either oligomerisation was not a factor, or that the K_d of any oligomeric species was below ~ 25 nM. Tyler Harmon produced predicted per residue secondary structure profiles for each of the sequences (Figure 6.3). The overall levels of predicted helicity agreed well with the CD data – BIK was predicted to be the most disordered, while PUMA was estimated to be the most helical. Predictions of the β-sheet content were also produced: PUMA, BIM and the BIM-swap showed < 2% β content, whereas both BIK and the BIK-swap demonstrated 7% and 6%, respectively. Therefore, although the β-sheet content of the BIK-swap was predicted to be higher than either PUMA or BIM, the prediction does not match the enhanced β content in the CD spectrum of the BIK-swap compared to BIK.

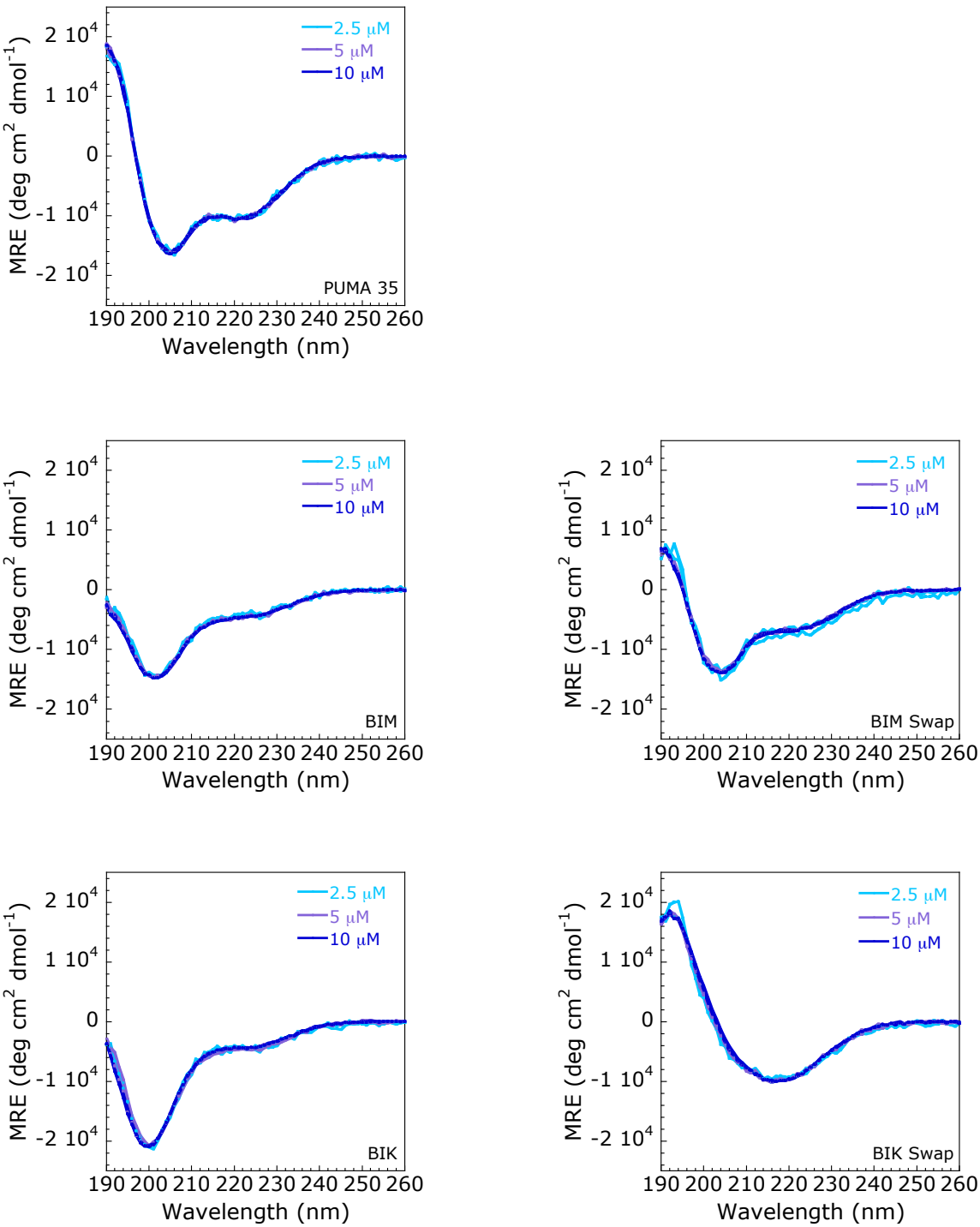


Figure 6.2 CD Spectra for PUMA, BIM, BIK and PUMA Core Swaps.
Concentration dependence for the CD spectra of PUMA, BIM, BIK and the PUMA core swaps. For the core swaps, all 19 PUMA residues that contact MCL-1 were kept in the same position. The residues from BIM (BIM-swap) or BIK (BIK-swap) were swapped into the position of the non-contacting residues. All proteins displayed a spectrum that was independent of the protein concentration. Each concentration was prepared in duplicate and both spectra are shown.

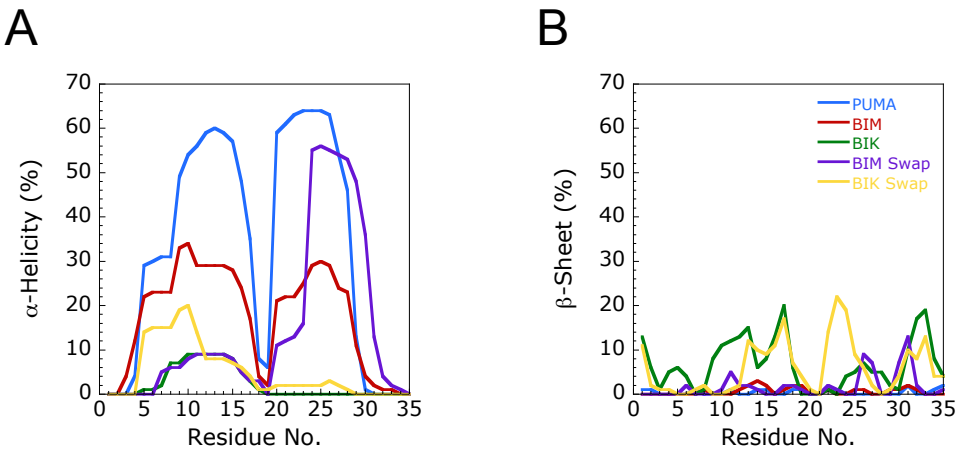


Figure 6.3 Predicted Secondary Structure Profiles for PUMA, BIM, BIK and Core Swaps.
Using ABSINTH (Vitalis and Pappu, 2010), Tyler Harmon predicted the secondary structure profiles of the peptide sequences. The mean α-helicity across the entire sequence was 33% for PUMA, 17% for BIM, 2% for BIK, 15% for the BIM-swap and 5% for the BIK-swap. The mean β-sheet content across the entire sequence was 0% for PUMA, 1% for BIM, 7% for BIK, 2% for the BIM-swap and 6% for the BIK-swap.

To determine if the altered residual structure of the peptides impacted the structure when bound to MCL-1, CD was performed on a mixture of the 2 proteins. When mixed with MCL-1, an increase in total helical content was observed for all peptides (Figure 6.4). MCL-1 was therefore able to interact with all of the peptides. Upon binding, the spectrum for each of the peptide:MCL-1 mixtures was similar, suggesting that each peptide formed a bound complex that was comparable to PUMA:MCL-1 (Figure 5.8).

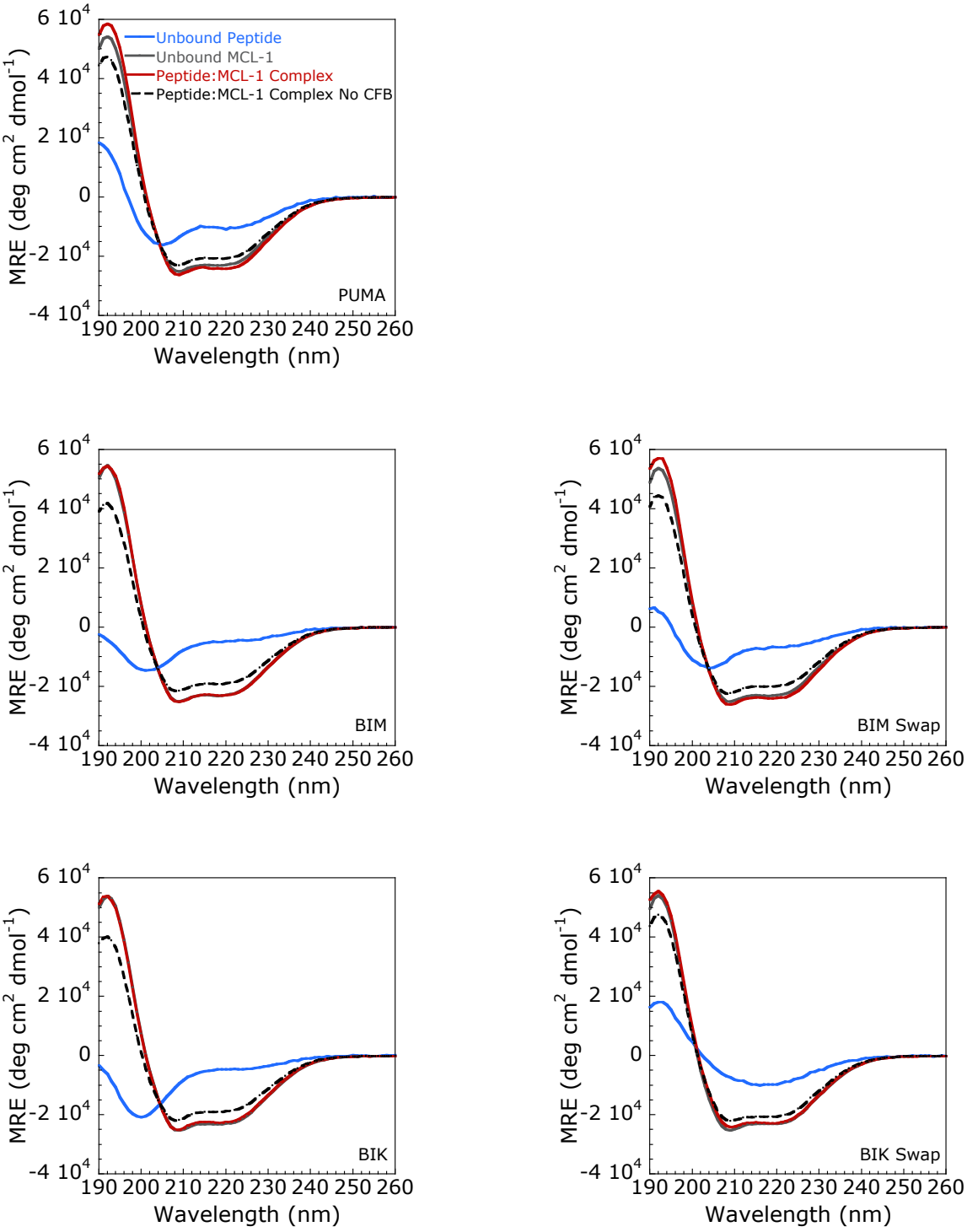


Figure 6.4 Coupled Folding and Binding of PUMA, BIM, BIK and Core Swaps Probed by CD.

A gain in α-helical structure was observed upon mixing each peptide with MCL-1. The expected spectrum (black, dashed line), if no folding upon binding occurred (no CFB), was calculated from the mean of the individual spectra.

6.1.2 Equilibrium Binding Affinity

The interaction between the 35-mer PUMA peptide and MCL-1 is reported as having a K_d of 0.18 ± 0.02 nM (Rogers *et al*., 2014a). Conventional methods of measuring affinity, such as isothermal calorimetry (ITC), are unsuitable for such tight complexes. An alternative method to measure the affinity of tight complexes is to outcompete another tight binding ligand. For example, a complex between a fluorescent dye-labelled peptide and MCL-1 can be outcompeted by addition of unlabelled PUMA. The concentration of unlabelled PUMA required to outcompete the labelled peptide is dependent on the affinity of MCL-1 for both the labelled peptide and the unlabelled PUMA. Therefore, if the affinity of the labelled peptide is known, the affinity of the unlabelled peptide can be determined (Nikolovska-Coleska *et al.*, 2004).

To perform the out-competition experiment, a suitable fluorescent dye-labelled peptide is required. A PUMA double mutant (A139G A150G) had been shown previously to bind MCL-1 with a K_d 1.7 ± 0.2 nM. This K_d just falls within the range that is measurable using fluorescence experiments. Higher affinity interactions would require sub-nM concentrations of peptide and MCL-1. At these low concentrations, the sticking of proteins to plastic and glassware can significantly alter the concentration, reducing the accuracy of the technique.

A 35-mer PUMA A139G A150G peptide with a fluorescence TAMRA dye attached to the N-terminus was purchased. Due to the importance of having an accurate K_d, both a kinetic and thermodynamic analysis of the interaction with MCL-1 was performed (Figure 6.5). A K_d of 1.7 ± 0.2 nM was obtained from equilibrium binding experiments, which was identical to the K_d obtained for the unlabelled peptide (Rogers *et al*., 2014a). The affinity was confirmed by taking the ratio of k_{off} over k_{on}, which gave a K_d of 2.06 ± 0.01 nM.

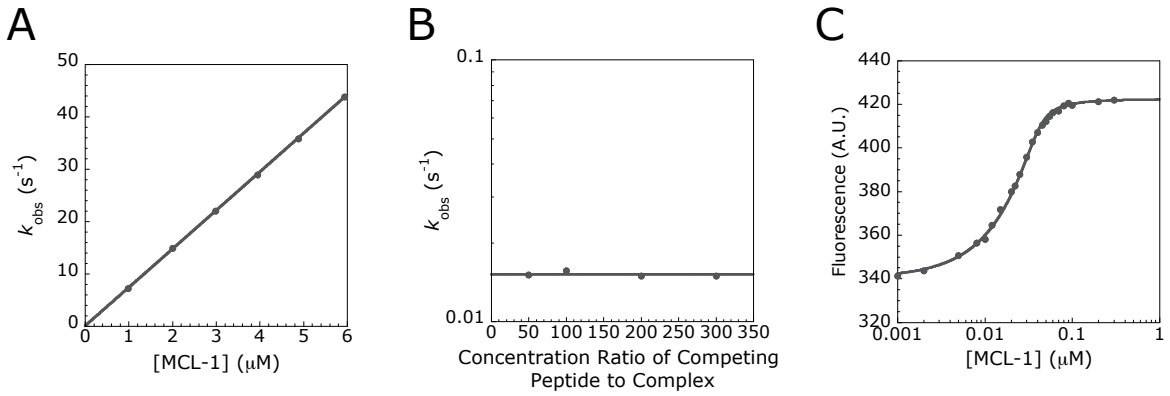


Figure 6.5 Kinetic and Thermodynamic Characterisation of PUMA A139G A150G Binding to MCL-1.
(A) 50 nM of dye-labelled A139G A150G PUMA was rapidly mixed with various excess concentrations of MCL-1. Changes in fluorescence intensity were fit to a single exponential function to extract k_{obs}. The gradient of the straight-line fit gave an association rate constant of 7.35 ± 0.03 $\mu M^{-1}s^{-1}$. (B) A pre-formed complex of dye-labelled A139G A150G PUMA and MCL-1 was mixed with various excess concentrations of unlabelled WT PUMA. Changes in fluorescence intensity were fit to a single exponential function to extract k_{obs}. The mean of the concentration independent k_{obs} gave a k_{off} of 0.01518 ± 0.00002 s^{-1}. (C) Various concentrations of MCL-1 were incubated with 300 nM of dye-labelled A139G A150G PUMA. A K_d of 1.7 ± 0.2 nM was extracted by fitting to Equation 3.15. Errors represent the error of the fit, except for the k_{off}, where the error represents the standard error of the mean.

Once the affinity of the dye-labelled peptide was determined, competition binding experiments were performed (Figure 6.6). The midpoint of the titration curve was used to calculate an equilibrium constant. As the value is calculated from a competition (inhibition) experiment, this is referred to as the K_i. PUMA (0.30 nM K_i), BIM (0.26 nM K_i) and the BIM-swap (0.30 nM K_i) all bound with similar affinity. BIK and the BIK-swap, on the other hand, had a K_i of 29 nM and 3 nM, respectively. Therefore, while swapping the non-contacting residues of BIM into the PUMA sequence appeared to have little effect, swapping in the residues from BIK reduced the affinity by more than an order of magnitude. Note that the error associated with the assay was not determined, as the experiment was performed once for each peptide.

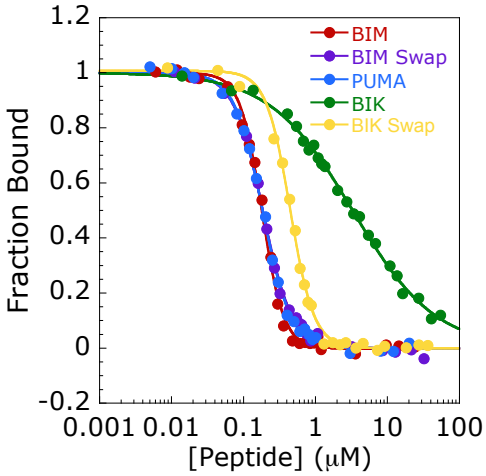


Figure 6.6 Competition Binding Experiments for PUMA, BIM, BIK and the Core Swaps.
Various concentrations of unlabelled PUMA, BIM, BIK, BIM core swap and BIK core swap were incubated with 300 nM of both dye-labelled A139G A150G PUMA and MCL-1. As the concentration of competition unlabelled peptide was increased, the amount of dye-labelled PUMA bound to MCL-1 decreased.

6.1.3 Change in Affinity is not due to a Difference in k_{on}

Association kinetics experiments were performed to determine if the altered affinity for the BIK core swap was due to a difference in the rate of association to MCL-1. Each peptide was rapidly mixed with MCL-1 using stopped-flow. In these experiments the proteins were mixed at near equimolar concentrations. The change in intrinsic tryptophan fluorescence with time was monitored and fit to a reversible bimolecular association model (Equation 2.26), which allows both the k_{on} and the K_d to be extracted. PUMA, BIM and the BIM-swap all displayed relatively concentration independent amplitude changes (Figure 6.7). Whereas, both BIK and the BIK-swap displayed an increase in signal amplitude as their concentrations were increased, consistent with a lower affinity for MCL-1 compared to PUMA, BIM and the BIM-swap.

At low concentrations of peptide, the bimolecular reversible association model appeared to deviate from the data (Figure 6.7). This was especially evident for both the BIM-swap and the BIK-swap peptides. One explanation for this deviation could be that the peptides did not display 2-state behaviour. However, previous

experiments showed that, at least for PUMA, the reaction occurred in a 2-state manner (Rogers *et al*., 2014a, 2014b; Rogers *et al*., 2013).

Fitting the traces individually rather than globally lead to an improved fit at low concentrations for the BIK-swap. However, for the higher concentrations, the K_d did not converge and either reported an unphysical negative value, or reached an imposed limit. This can occur when the protein concentrations are far above the K_d. Unfortunately, this was the case for BIM, PUMA and the BIM-swap, which all bound too tightly to be able to determine the K_d using this method.

On the other hand, global fitting reported a K_d of 12.1 ± 0.1 nM for the BIK-swap and 44.8 ± 0.1 nM for BIK. These K_d values are consistent with the both the equilibrium data (K_i = 3 nM for BIK-swap and 29 nM for BIK) and the increase in amplitude observed when the concentration of peptide was increased in the kinetic experiment (Figure 6.7). Furthermore, both the global fit and individual fitting method gave similar k_{on} values of between 14 and 23 $\mu M^{-1}s^{-1}$ for the BIK-swap. Fitting the data globally therefore provided an estimate of the K_d, without altering the estimate of the k_{on}. Hence, despite resulting in a poorer fit for some traces, the global fit method was used to obtaining an estimate of the K_d for BIK and the BIK-swaps. Sharing the parameters by globally analysing the data should also improve the accuracy of the k_{on} estimate.

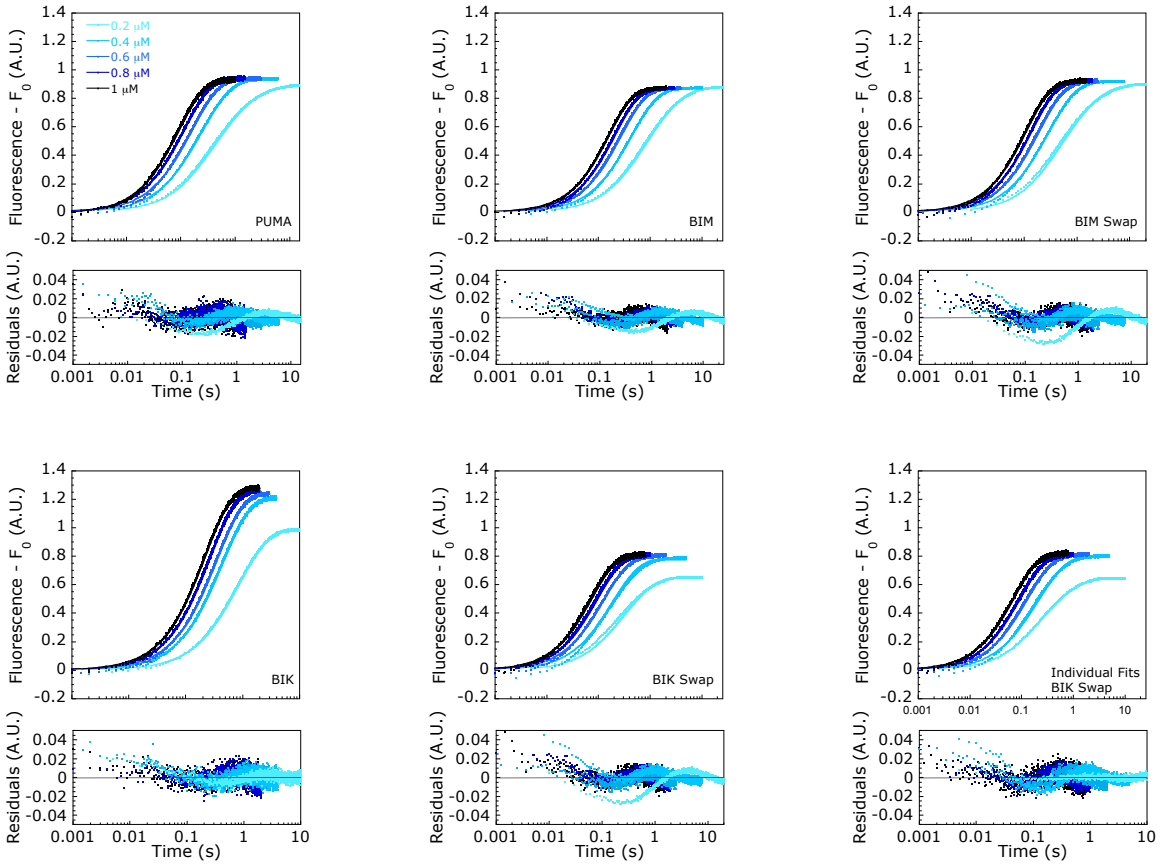


Figure 6.7 Kinetic Traces for PUMA, BIM, BIK and the Core Swaps
Various concentrations of each peptide were mixed with 0.2 μM of MCL-1. Changes in intrinsic tryptophan fluorescence with time were monitored and globally fit to a reversible bimolecular association model (Equation 2.26). Individually fitting the BIK-swap data (bottom right) resulted in improved residuals compared to the global analysis (bottom middle); however, the K_d could not be extracted. Data represent the mean of 15 – 20 individual traces.

The association rate constants for all the peptides were within 3-fold of each other (Figure 6.8). With a k_{on} of 3.5 ± 0.02 $\mu M^{-1}s^{-1}$, BIK displayed the slowest association rate constant. Although not yet examined, it could be hypothesised that this was due to the loss of 2 charged residues, reducing the electrostatic steering. At the other end of the spectrum, the BIK-swap had the fastest k_{on} of 14.45 ± 0.02 $\mu M^{-1}s^{-1}$. This was surprising, given that the BIK-swap peptide had β-sheet-like residual structure and contained the non-contacting residues of the slowest associating peptide (BIK). Similar effects were observed for the BIM-swap peptide, which associated almost twice as fast as BIM. Assuming that the peptides displayed 2-state behaviour, the difference in affinity for the BIK-swap was solely due to an

increase in the k_{off}. In contrast, the affinity of BIK was reduced due to a reduction in k_{on} and an increase in k_{off}.

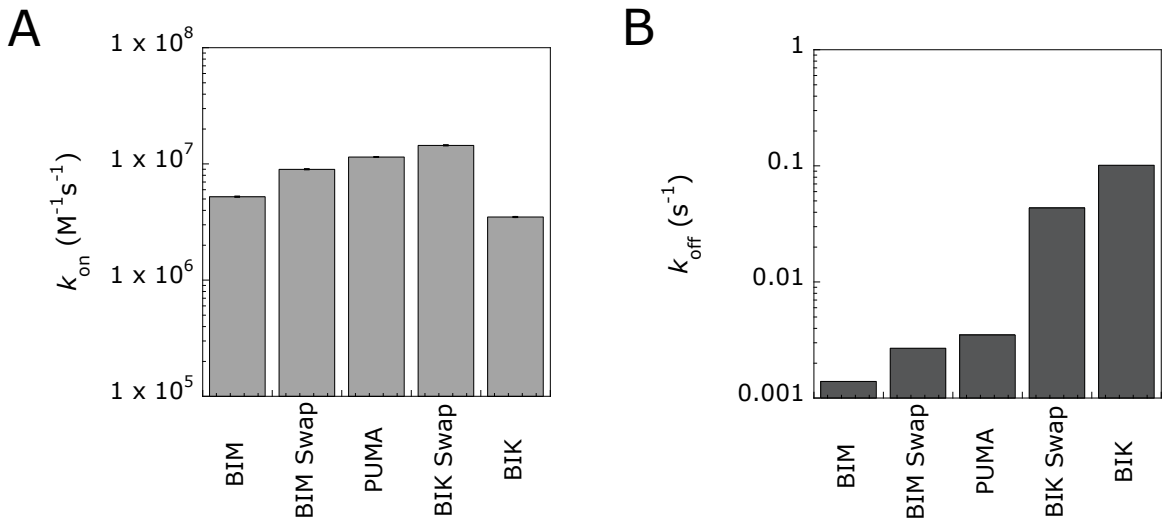


Figure 6.8 Rate Constants for PUMA, BIM, BIK and the Core Swaps.
(A) Association rate constants were determined as 5.23 ± 0.04 $\mu M^{-1}s^{-1}$ for BIM, 9.01 ± 0.08 $\mu M^{-1}s^{-1}$ for the BIM-swap, 11.53 ± 0.01 $\mu M^{-1}s^{-1}$ for PUMA, 14.45 ± 0.02 $\mu M^{-1}s^{-1}$ for the BIK-swap and 3.50 ± 0.02 $\mu M^{-1}s^{-1}$ for BIK. Errors and error bars represent the error of the fit. (B) Dissociation rate constants could be estimated from multiplying K_i by k_{on}.

6.2 Discussion

The influence of non-contacting PUMA residues on the binding affinity for MCL-1 was probed by swapping non-contacting residues from other BH3-only proteins into PUMA. PUMA, BIM and BIK all undergo coupled folding and binding upon interacting with MCL-1. In isolation, BIM and BIK displayed a similar, predominantly disordered structure, whereas PUMA showed an element of residual helical structure. Surprisingly, when the non-contacting residues of BIK were swapped into the same relative position in PUMA, the BIK core swap peptide displayed residual β-sheet-like structure. Swapping the non-contacting residues of BIM into PUMA also influenced the residual structure, reducing helicity from 29% in PUMA to 18% in the BIM-swap. Despite the change in residual structure, PUMA, the BIM-swap and the BIK-swap all bound with similar association rate constants, which is consistent with the reported induced fit mechanism of binding for PUMA (Rogers *et al*., 2014b).

Although this work is at a relatively preliminary stage, it is interesting to note that the affinity of the BIK-swap was reduced by an order of magnitude compared to PUMA. Combined with the data from the conserved shuffles (Chapter 5.3.3), this suggests that non-contacting residues have an important role to play in determining affinity. This hypothesis is consistent with recently published data, showing that when mutations that enhance selectivity of PUMA for A1 are put into the context of BIM, they reduce the affinity by more than 3 orders of magnitude (Jenson *et al.*, 2017). Thus, the mutations only enhance selectivity in the context of the PUMA sequence. When placed into another, tight binding BH3-only protein, they have the opposite effect. Furthermore, the importance of the background protein sequence has been recently highlighted in antibody binding studies (Boyer *et al.*, 2016). Typically, when producing antibodies, a few residues are varied and the remaining framework is left constant. Yet, Boyer *et al* (2016) found that the ability to produce selective binders was dependent on the background context of the variable residues.

It is too early to conclude why the non-contacting residues appear to influence binding affinity; however, a few reasons may be speculated. Firstly, there is an enthalpic argument. The specific sequence of non-contacting residues may provide a framework that positions the contacting residues in specific places. When the sequence is altered, by shuffling or swapping, the position of the contacting residues could slightly shift, reducing the interaction energy. Secondly, the other component of Gibbs free energy is entropy. Non-contacting residues can influence the residual structure of the peptide, which could affect the entropic cost of folding and binding. In other words, for a peptide that forms a helix upon binding, reducing the residual helicity will require a greater degree of folding upon binding. Thirdly, there are other components to entropy other than just considering the structural distance between the free, unfolded and bound, folded peptide. For example, a certain sequence may provide significantly more side chain steric clashes when folding, limiting the position of the side chains and increasing the entopic cost of folding. This idea is in line with data showing that phi and psi angles can be influenced by neighbouring residues (Creamer and Rose, 1992; Horovitz *et al.*, 1992; Srinivasan and Rose, 1999). During all this speculation, one point of note should be made from the current data. In opposition to the second view, the BIM-swap bound with >5-fold higher affinity than the conserved shuffles, despite having

18% residual helicity compared to 22%, 26% and 23% for conserved shuffle 1, 2 and 3, respectively.

In summary, while this work is still to be completed, it may prove important for drug design, where the focus typically lies on the contacting residues, with little heed paid to the rest of the sequence. Engineering the non-contacting could enhance binding affinity significantly.

Chapter 7

BCL-2 Family

Equilibrium Chemical denaturation experiments for A1 were collected and analysed with Carolina Mendonça. Basile Wicky and Tristan Kwan collected all data related to the proteins, BAK and BAX.

7.1 The Role of BCL-2 Family Proteins in Apoptosis

Evolution dictates that only the fittest, best adapted organisms survive and reproduce, passing on their characteristics to the next generation. For unicellular organisms, this puts the emphasis on the individual cell's survival. Benefits can still be gained by working with neighbouring organisms; for example, bacterial biofilm formation aids the survival of the whole population (Donlan, 2001; Kostakioti *et al.*, 2013). However, through competing for nutrients, an individual bacterium still prioritises its own survival at the expense of its neighbours. A multicellular organism does not carry the same luxury: any cell that is unhealthy or energetically costly decreases the organism's fitness, reducing the likelihood that it will survive through evolution. Counterbalancing this disadvantage is the ability to separate and optimise specific functions through differentiating cells into tissues and organs. Cell differentiation is a relatively irreversible process under standard physiological conditions. This creates an issue during an organism's development, as cells with different functions are required at specific times - a differentiated cell or tissue may be crucial for early stages of development, but not needed once the organism reaches maturity. As maintaining these unnecessary cells requires nutrients, a significant evolutionary advantage can be gained by safely disposing of cells that provide no functional benefit. Accordingly, apoptosis, the process of controlled cell death, has evolved to allow organisms to safely dispose of unhealthy or unwanted cells (Ameisen, 2002). For example, during development, 148 out of 1179 cells in the round worm, *Caenorhabditis elegans*, can be visually observed to undergo apoptosis (Ellis and Horvitz, 1986).

There are two main mechanisms by which apoptosis occurs (Elmore, 2007). In the first mechanism, tumour necrosis factor (TNF) cytokines bind to TNF receptors on the extracelluar surface of the plasma membrane. Ligand binding triggers a series of intracellular interactions that result in apoptosis through the activation of caspases (Li and Yuan, 2008). As this mechanism is initiated by the binding of an extracellular ligand, this is known as the extrinsic pathway. The second pathway is activated by intracellular signals, and is known as the intrinsic pathway. Intracellular initiators include the absence of growth factors, hormones or cytokines; or the presence of radiation damage, toxins, hypo/hyperthermia, viral infections and free radicals (Elmore, 2007). These stimuli activate BH3-only family members, which interact with BCL-2 family proteins. The interaction between BH3-only proteins and the BCL-2 family elicits a response that leads to oligomerisation and pore formation in the outer mitochondrial membrane (Figure 7.1). Upon pore formation, various factors (e.g. cytochrome-c) are released from the mitochondria, and apoptosis is initiated.

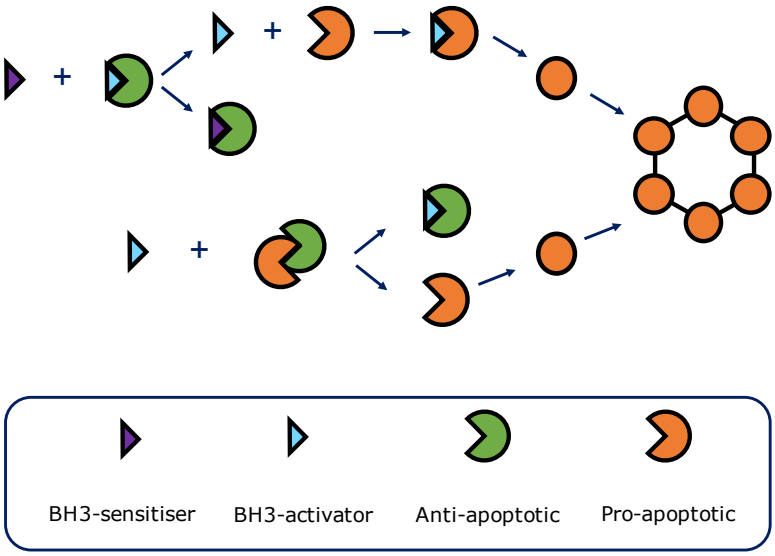


Figure 7.1 Mechanisms of BCL-2 Family Pore Formation by BH3-Only Proteins.
Proteins in the BCL-2 family are split into 2 functional classes based on whether they promote apoptosis through forming pores (pro-apoptotic), or inhibit apoptosis (anti-apoptotic). BH3-only proteins can also be separated into 2 categories – those that can bind directly to either the pro-apoptotic or anti-apoptotic members (activators) or those that can only bind to the anti-apoptotic proteins (sensitisers). Anti-apoptotic proteins sequester either BH3-activators (top scheme) or pro-apoptotic BCL-2 proteins (bottom scheme). BH3-only proteins compete for binding to the anti-apoptotic proteins, releasing the sequestration and promoting pore formation of the pro-apoptotic proteins (Czabotar et al., 2014; Shamas-din et al., 2013).

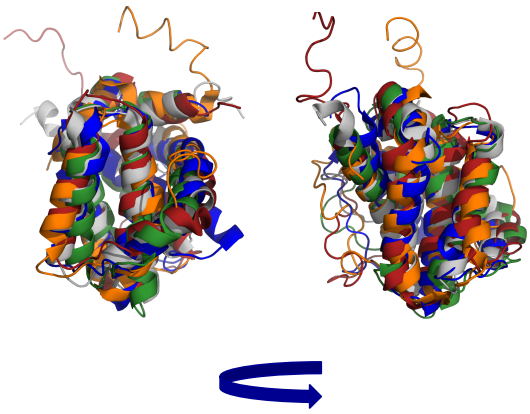
7.2 Members of the BCL-2 Family

There are 6 anti-apoptotic (BCL-2, MCL-1, A1, BCL-XL, BCL-W and BCL-B) and 3 pro-apoptotic members (BAK, BAX and BOK) of the BCL-2 family. For the purposes of this research, 3 of the anti-apoptotic and 2 of the pro-apoptotic proteins were selected for analysis. Mouse MCL-1 had been previously studied by Dr Joe Rogers. To allow comparison, mouse A1 and BCL-XL were chosen as the two additional anti-apoptotic proteins. Human BAK and BAX were utilised, as structures for the mouse versions of the pro-apoptotic proteins were not available.

Despite having opposing functions and a sequence identity of < 30%, BAK, BAX, MCL-1, A1 and BCL-XL all share a homologous structural fold (Figure 7.2). Helicies

2 – 5 and 8 form a hydrophobic grove, creating the interaction surface for BH3-only proteins.

A



B

		RMSD (Å)				
		MCL-1	A1	BCL-XL	BAK	BAX
Sequence Identity (%)	MCL-1		3.4	3.9	2.5	2.1
	A1	26		2.6	2.5	2.9
	BCL-XL	24	29		4.2	4.7
	BAK	26	24	26		3.4
	BAX	25	22	23	25	

Figure 7.2 Structure and Sequence Comparison for BCL-2 Family Proteins. (A) Structures for the BCL-2 proteins A1 (green; PDB code 2VOI), MCL-1 (grey; PDB Code 2ROC), BCL-XL (orange; PDB code 2M04), BAK (blue; PDB code 2M5B) and BAX (red; PDB code 1F16) were aligned using PyMOL (version 1.7.2.1, Schrödinger). For each protein, only the BCL-2 domain is shown (e.g. for MCL-1 the N-terminal disordered region was not included in the protein used to determine the structure). (B) Sequences for the proteins investigated in this work were aligned using Clutsal Omega (Goujon *et al.*, 2010; Sievers *et al.*, 2014) and the resultant sequence identities are shown. RMSD values were determined from the PyMOL structural alignments of the backbone atoms.

7.2.1 BH3-Only Proteins

BCL-2 like proteins can contain up to 4 BCL-2 homology (BH) motifs, with all members containing at least the BH1 and BH2 motifs. The BH3 motif, consisting of the sequence LXXXGDE (where X is any amino acid), is also present in the BH3-only proteins. BH3-only proteins are typically intrinsically disordered (Czabotar et al., 2014; Hinds et al., 2007; Rogers et al., 2013); although, BID is a notable folded exception to this generality (McDonnell *et al.*, 1999) (Figure 7.3). Through linking the extrinsic and intrinsic apoptosis pathways, BID also displays a unique functional role. After activation of the extrinsic pathway, caspase-8 cleaves cytosolic BID,

forming truncated BID (t-BID). t-BID translocates to the mitochondrial membrane, where it promotes pore formation of pro-apoptotic BCL-2 proteins (Li *et al.*, 1998).

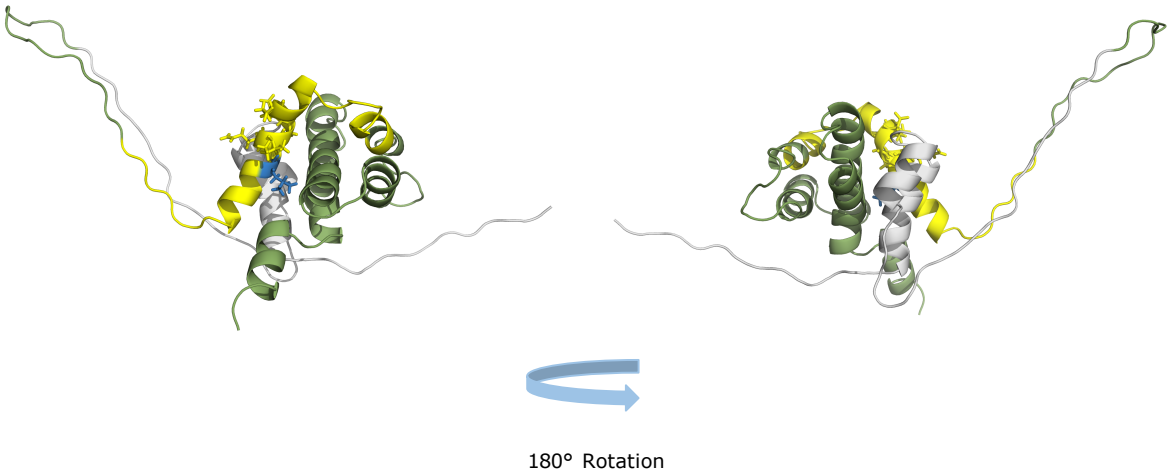


Figure 7.3 Folded Structure of Full Length BID.
The BH3 motif of mouse BID (sticks) forms part of the fold. Upon interacting with a BCL-2 member, the leucine (L90) in the BH3 motif becomes buried in the hydrophobic grove. In folded BID, this leucine (blue) is buried in the core of the protein, and is unavailable for binding. To allow interaction, BID may therefore have to unfold before binding. In this work, a 35-residue segment of BID was investigated (yellow). During its activation, 60 N-terminal residues of full length BID (grey) are cleaved by caspases to form truncated BID (green and yellow; BID p15). Other cleavage products of BID also occur, with residues 1 – 75 (BID p13) or 1 – 98 (BID p11) cleaved (Wei *et al.*, 2000). Structures were generated using PyMOL (version 1.7.2.1, Schrödinger) and are based on PDB code 1DDB.

7.3 Results

7.3.1 Chemical Denaturant Induced Oligomerisation

For more than a century and a quarter, the effect of ions on protein stability has been noted (Hofmeister, 1888). While the molecular origin of the effect is still not well understood, it has been suggested that it is due to interactions of salt ions with the protein backbone and charged amino acids (Okur *et al.*, 2017). For ions that destabilise 3-dimensional folds, more interaction sites are available when the folded protein chain is exposed to the solvent, resulting in stabilisation of the denatured, unfolded state. Altering the stability ($\Delta G_{D\text{-}N}$) of a protein by titrating destabilising ions provides a powerful method to unfold proteins and estimate their stability in the absence of denaturant (Myers *et al.*, 1995). Changes in protein structure can be monitored by following the CD spectra. Alternatively, fluorescence from intrinsic tryptophan residues can be used as a proxy for protein structure: exposure of a buried tryptophan to solvent during unfolding typically alters the fluorescence intensity and/or the peak emission wavelength.

When exposed to the destabilising chemical denaturants, urea or guanidine hydrochloride (GdmCl), members of the BCL-2 family show either one (MCL-1) or two (A1, BCL-XL, BAK and BAX) transitions (Figure 7.4). Biphasic unfolding curves suggests that a stable species is formed along the unfolding pathway. The fluorescence profile for the biphasic curves appeared very similar – the first transition was associated with a large increase in fluorescence intensity, the second displayed a smaller, reduction in fluorescence intensity. A shift in the emission wavelength from ~ 320 nm to ~ 345 nm was observed for the first transition. The second transition was associated with a further shift, reaching approximately 360 nm (Figure 7.5). CD indicated that the first transition was associated with a loss of around 30 – 40% of the helicity. The remaining helicity was lost during the second transition.

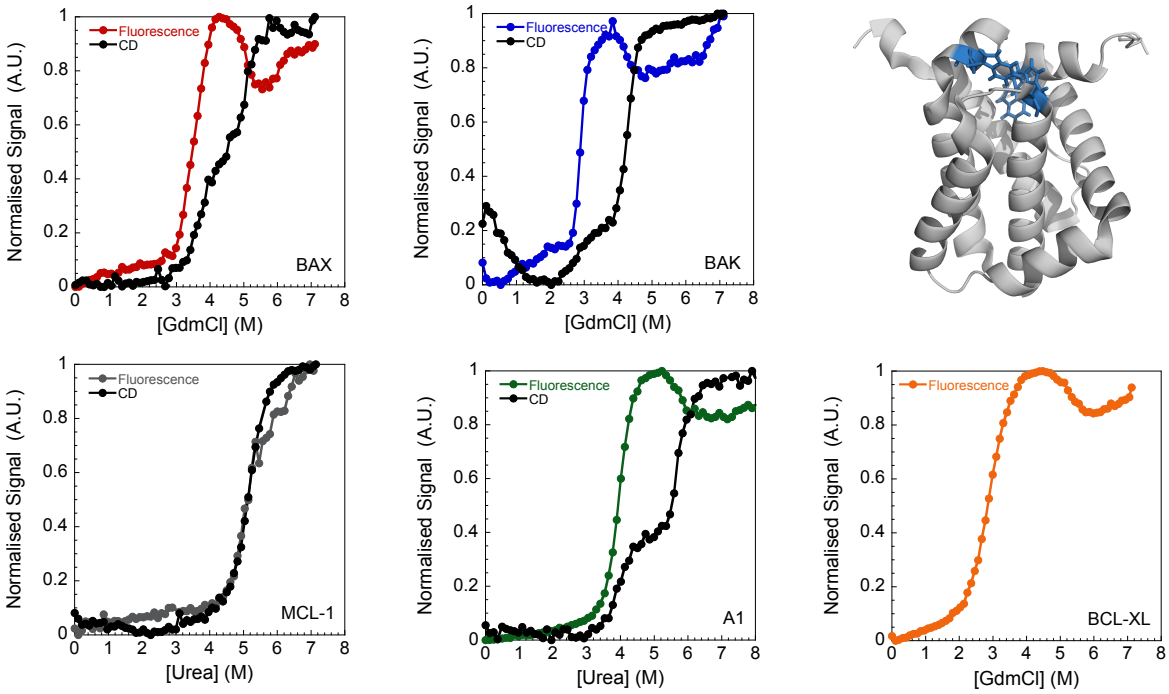


Figure 7.4 Equilibrium Chemical Denaturation Curves for BCL-2 like Proteins.

For unfolding experiments, approximately 1 μM of each protein was incubated with various concentrations of a chemical denaturant (urea or guanidine hydrochloride). Different proteins displayed varying absolute fluorescence intensity and CD values. Therefore, to aid comparison, the recorded signal was normalised. The CD signal represents the normalised value at 222 nm. Addition of denaturant is associated with a loss of total helical structure - the buffer subtracted mdeg values transition from a negative value at 0 M denaturant, to close to 0 mdeg at 8 M. Each examined BCL-2 like protein contains 3 conserved tryptophan residues (shown as blue sticks in the structure). Additional tryptophan and tyrosine residues are also present in various, non-conserved positions. CD data was not collected for BCL-XL, as the buffer contained both DTT and NaCl, which absorbed light and caused the detector voltage to go over range. The structure was generated using PyMOL (version 1.7.2.1, Schrödinger) and is based on PDB code 2ROC.

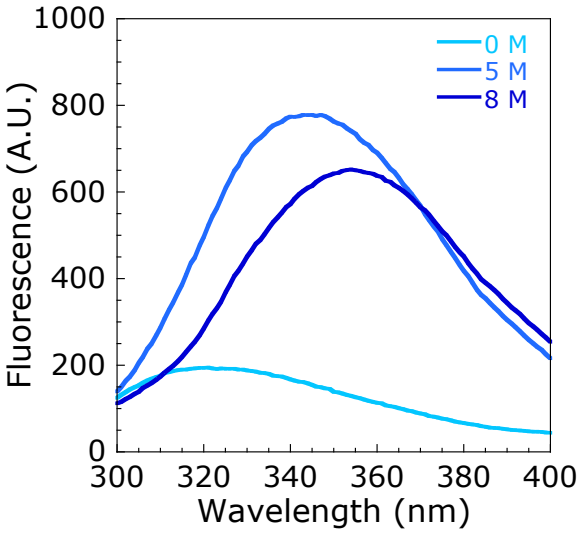


Figure 7.5 Shift in A1 Peak Emission Wavelength for Each Transition.
The fluorescence spectra for A1 obtained under 3 different concentrations of urea. Each concentration relates to the beginning of the curve (0 M), the peak of the first transition (5 M) and the end of the curve (8 M). For clarity, the spectra obtained between these concentrations are not shown. A shift in emission wavelength from ~ 320 nm to ~ 345 nm to ~ 360 nm was observed as the concentration of urea was increased. The maximum emission wavelength of a buried and solvent exposed tryptophan is typically ~ 320 nm and ~ 360 nm, respectively.

One explanation for biphasic equilibrium unfolding curves could be the presence of a partially unfolded intermediate, that, due to having a solvent accessible surface area between the folded and unfolded state, becomes the most stable species at a given concentration of denaturant. Alternatively, either the formation or dissociation of an oligomeric species could explain the presence of 2 transitions. Experimentally, these two options can be distinguished by repeating the chemical denaturant titrations at different protein concentrations. If the biphasic behaviour is due to an intermediate, then the normalised curves should overlay as the process is unimolecular and, consequently, independent of the protein concentration. On the other hand, the presence of an oligomeric species is at least a bimolecular reaction. Any transition relating to the formation or dissociation of an oligomer would therefore be dependent on the protein concentration.

Consistent with a unimolecular unfolding process, no change in the single MCL-1 transition was observed upon increasing the protein concentration (Figure 7.6). Anti-apoptotic A1 and pro-apoptotic BAX both showed biphasic behaviour at 1 μM.

When the protein concentration was reduced, the first and second transitions of A1 shifted. As the transitions were concentration dependent, the biphasic nature of the curves was likely due to oligomers. Analysing the direction of the transition shift can provide insight into what process the individual transitions relate to. For lower concentrations of A1, the transition at low urea concentrations became less steep and shifted to the right, while the transition at high urea concentrations shifted to the left and became steeper. This pattern is consistent with the first transition indicating monomer to oligomer formation, with the second transition representing dissociation of the oligomer into unfolding monomer – at higher protein concentrations, oligomer formation would be more favourable, leading to a steeper low [urea] transition. The high [urea] transition would shift to the right, as the oligomer would be more stable, requiring greater amounts of denaturant to dissociate it. In contrast, if the low [urea] transition was an oligomer dissociation event, then it should shift to the left with decreasing protein concentration. A similar pattern of concentration dependent transitions was also potentially observed for BAX, although this needs to be repeated at an additional concentration.

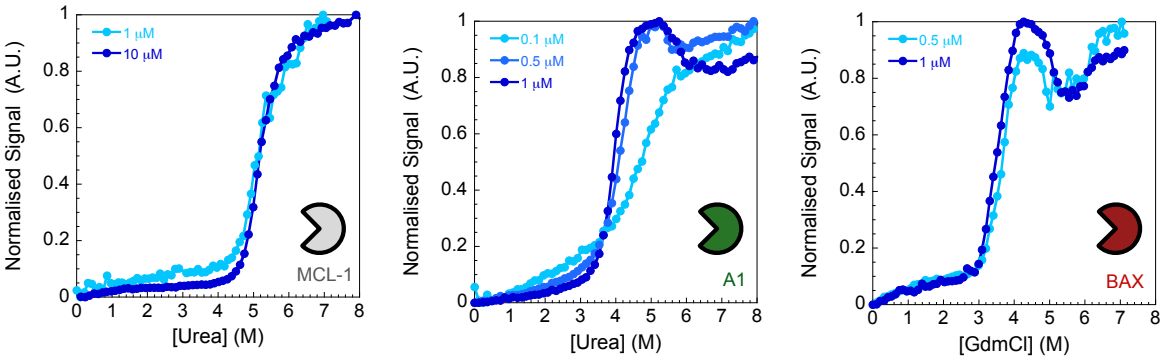


Figure 7.6 Biphasic Behaviour Due to Formation of an Oligomeric Species. Equilibrium chemical denaturation titrations were performed at various concentrations of MCL-1, A1 and BAX. To aid comparison, the fluorescence intensity signal was normalised. MCL-1 displayed a single transition at both 1 and 10 μM. The biphasic behaviour of both A1 and BAX was concentration dependent.

A monophasic unfolding curve was obtained at a low (0.1 μM) concentration of A1 (Figure 7.6). Either the 2 transitions had merged and could no longer be distinguished, or there was no oligomer present and the change in fluorescence was reporting solely on the unfolding of monomeric A1.

7.4 Discussion

To trigger controlled cell death, pro-apoptotic BAK and BAX oligomerise, forming pores in the mitochondrial membrane and committing the cell to apoptosis (Chi *et al.*, 2014). The mechanism for the oligomerisation has been a subject of much debate (Bogner *et al.*, 2010; Dewson *et al.*, 2008; Shih *et al.*, 2017). A possible mechanism is the insertion of the BH3 motif of one BCL-2 like protein into the hydrophobic grove of another, which has been shown to result in homodimers (Dewson *et al.*, 2008). In this model, higher order oligomers are then formed via interactions between α6 of each protein (Dewson *et al.*, 2009; Shih *et al.*, 2017).

When present, the BH3 motif of BCL-2 like proteins is typically buried and inaccessible for binding. Some form of structural rearrangement, or unfolding of the protein would therefore be required to allow dimerisation. This could explain why adding a chemical denaturant, which destabilises the folded state of the protein, can result in the formation of oligomers (Figure 7.6). Consistent with this hypothesis of unfolding and binding, the CD data indicated that 30 – 40 % of the total helicity was lost upon forming the oligomer (Figure 7.4). Although structural data are required to make firm statements about the oligomeric state observed in the chemical denaturant assays, the data are consistent with the proposed models for the formation of functional, pore-forming oligomers. Consequently, it is tempting to speculate that the oligomers observed in the denaturant assays are related to either the dimeric or higher order, functional species.

7.4.1 Evolution of the Oligomeric Species

Of the proteins investigated, only BAK and BAX are reported to form pores in the outer mitochondrial membrane (Czabotar *et al.*, 2014). However, oligomerisation in the presence of chemical denaturant was observed for both pro-apoptotic BAK and BAX, and the anti-apoptotic BCL-2 like proteins. Not all proteins oligomerised to the same extent – MCL-1 appeared to remain monomeric at all concentrations examined, whereas A1 showed a concentration dependent formation of oligomers in the high nM range. The pro-apoptotic and anti-apoptotic proteins are structurally homologous, but share little sequence identity (Figure 7.2). As the pro- and anti-apoptotic proteins are so structurally homologous, it is interesting to speculate which came first. Did the ability to form pores evolve from the anti-apoptotic proteins, creating the pro-apoptotic proteins? Or were the less oligomerisation-

prone, anti-apoptotic proteins evolved from the pore forming pro-apoptotic proteins?

In an attempt to answer these questions, it would potentially be extremely insightful to reconstruct the evolutionary ancestor of the BCL-2 like proteins (Figure 7.7). As well as providing information about the functional evolution of the family, ancestral reconstruction may also indicate what underlies the differing oligomerisation propensities. For example, which residues are key for forming intermolecular contacts. Work on reconstructing the evolutionary ancestor of the BCL-2 like proteins has begun.

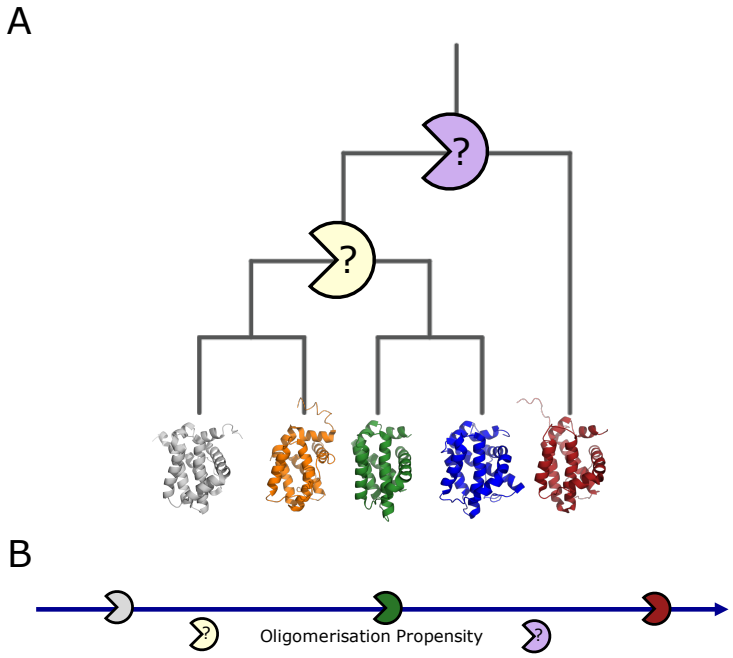


Figure 7.7 Where do the Evolutionary Ancestors fall on the Oligomerisation Propensity Scale?
(A) To reconstruct the evolutionary ancestor, modern-day protein sequences are aligned. A statistical model is then used to track the most likely paths between species, and this is used to infer the sequence of the ancestral protein. Both the ancient protein and evolutionary intermediates can, in theory, be reconstructed (Thornton, 2004). (B) A spectrum of oligomerisation propensity was observed in the chemical denaturant experiments. Where do the evolutionary ancestors fall on this scale?

Chapter 8

Encoding of Coupled Folding and Binding Information

For a protein, successfully achieving the correct functional fold poses quite a challenge. Each phi and psi angle has the potential to explore a large variety of conformations when unfolded, but must be constrained into a relatively fixed position in the folded state. To avoid the large potential of forming an incorrect structure, proteins follow folding pathways. As demonstrated by Anfinsen's experiments, all of the information required to correctly fold a protein can be contained within the amino acid chain (Anfinsen, 1973; Anfinsen *et al.*, 1961). Thus, proteins have evolved to act as their own instruction manuals.

As seen in the previous chapters, a subset of intrinsically disordered proteins can fold upon binding to a partner macromolecule. In these cases, the folding reaction of the IDP occurs in the presence of another protein chain. This raises an interesting question – where is the information encoded? Is it contained within the IDP? Or does the partner template the folding and binding?

Many pieces of information can be encoded within the amino acid sequence – the folded structure; the stability; the rates of folding and unfolding; and the function. In terms of encoding the folding reaction, one aspect that is particularly intriguing is which residues are involved in committing the protein to folding. In other words, which amino acids form interactions at the reaction transition state. These residues, which guide protein folding through forming native-like interactions at the transition state, can be identified using a method known as φ-value analysis (see Chapter 1.6.1).

A handful of studies have also applied φ-value analysis to the study of IDP coupled folding and binding reactions (Dogan *et al*., 2013; Giri *et al*., 2013; Haq *et al*., 2012; Hill *et al*., 2014; Rogers *et al*., 2014a; Toto *et al*., 2016, 2014). In one of these studies, Rogers *et al*. (2014) investigated the folding and binding of the IDP,

PUMA to the folded protein, MCL-1. An early transition state is observed, with few interactions formed and little helical structure. Higher φ-values are present at the N-terminal end of PUMA, indicating that this region of the IDP is important in stabilising the transition state (Figure 8.1).

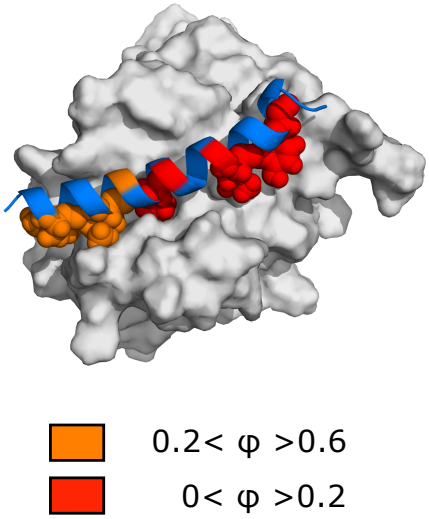


Figure 8.1 φ-Value Analysis for PUMA Binding and Folding to MCL-1
Higher φ-values are observed towards the N-terminus of PUMA, indicating a greater degree of native-like interactions at the transition state in this region. No fully native interactions are formed and the transition state occurs early along the reaction co-ordinate (generally low φ-values observed). The structure was generated using PyMOL (version 1.7.2.1, Schrödinger) and is based on PDB code 2ROC. Residues that are buried in the hydrophobic grove are shown as spheres. Data from Rogers *et al.* (2014).

As discussed in Chapter 7, the BCL-2 family contains a collection of intrinsically disordered BH3-only proteins that fold upon binding to structured, BCL-2-like proteins. The bound structures of the BH3-only:BCL-2-like complexes are highly homologous, yet each individual protein component shares <26% sequence identity (Figure 8.2). Based on Anfinsen's studies, as both the IDPs and the partner proteins have different primary sequences, they could contain different folding and binding instructions. Combined with the structural similarity of the bound complexes, this makes the BH3-only:BCL-2-like protein interaction network an ideal model system to investigate where the coupled folding and binding information is encoded. The structural similarity of the bound complex is an

important component to answering this question, as different residues have equivalent positions in each structure. Consequently, only the side chain of each residue has changed, with the encoded final fold remaining the same. Due to the importance of knowing the bound structure, only BH3-only:BCL-2-like protein interactions that had experimentally determined structures were considered.

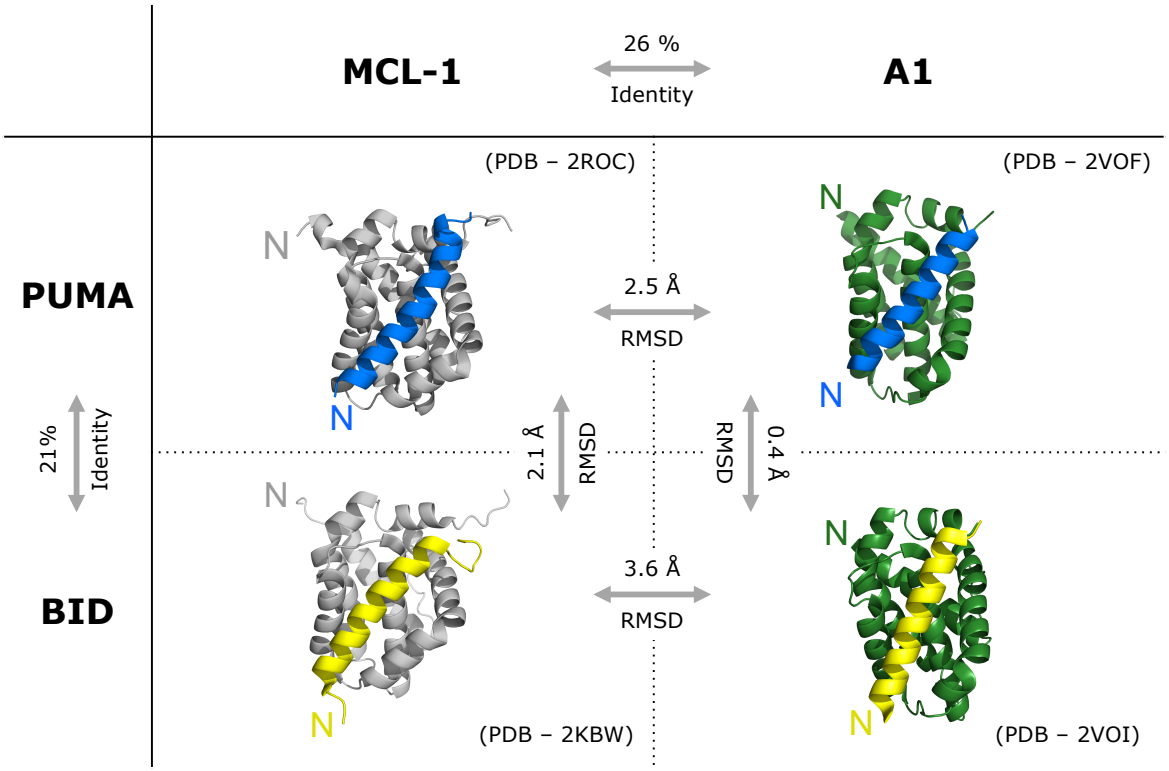


Figure 8.2 Structural and Sequence Comparison of BH3-Only:BCL-2-Like Complexes.
Sequences for the proteins investigated in this work were aligned using Clutsal Omega (Goujon *et al.*, 2010; Sievers *et al.*, 2014) and the resultant sequence identities are shown. RMSD values were determined from the PyMOL (version 1.7.2.1, Schrödinger) structural alignments of the backbone atoms. The N-terminus of each protein is indicated with an N.

To allow comparison with the published BH3-only:BCL-2-like transition state (Rogers *et al.*, 2014a), complexes that contained either PUMA or MCL-1 were sought. Analysis of the Protein Data Bank indicated that the BH3-only protein, BID or the BCL-2-like protein, A1 would be good choices. A1 had structures with both

BID and PUMA (Smits *et al.*, 2008), while BID had structures with both MCL-1 (Liu *et al.*, 2010) and A1 (Figure 8.2).

With the systems chosen, investigation of their coupled folding and binding reactions could begin. Specifically, the aim of this work was to determine the residues of the IDP that form interactions in the transition state. If the IDP was encoding the folding pathway, then the φ-values for PUMA should be the same when binding to either MCL-1 or A1. Whereas, if the partner protein contained the transition state information, the φ-values should be the same for PUMA and BID when binding to MCL-1.

8.1 Results

8.1.1 Purification of A1

Prior to the work described in this thesis, the expression and purification of A1 had not previously been attempted in the lab. A1 was expressed with a GST tag, allowing the protein to be pulled down on glutathione sepharose (GS). A TEV cleavage site was present between the GST tag and A1, allowing A1 to be cleaved off the resin (Figure 8.3). Ion exchange chromatography resulted in pure A1. As a final purification step, and to buffer exchange the protein, A1 was passed down a Superdex 75 gel filtration column. A1 eluted at a similar volume to MCL-1, another BCL-2 like protein of comparable size. The identity of A1 was then confirmed by mass spectrometry, which was performed by the Mass Spectrometry Service in the Department of Chemistry, Cambridge. Note that by SDS PAGE, A1 runs further than would be expected for a protein of its mass.

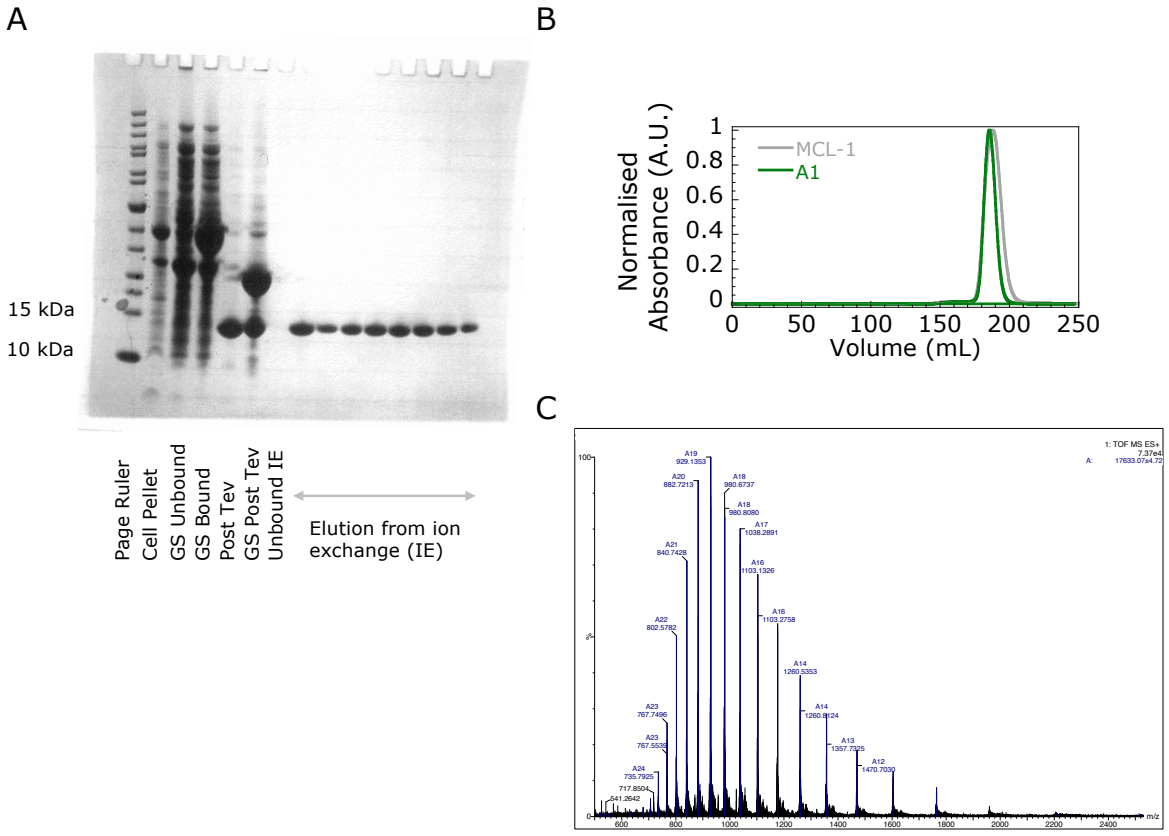


Figure 8.3 Purification of the BCL-2-Like Protein, A1.
(A) The GST-A1 fusion protein expressed in the soluble fraction. Addition of TEV protease resulted in full cleavage of A1 from the GST tag, releasing the protein from the glutathione sepharose (GS). Cleaved protein was purified using an anion exchange column, producing clean fractions according to SDS PAGE. A1 was buffer exchanged and further purified using a Superdex 75 gel filtration column. (B) The protein eluted at a similar volume to MCL-1. (C) Mass spectrometry provided a mass of 17633 ± 5 Da, which matched the expected mass of 17635 Da for the A1 construct.

8.1.2 Probing the Reaction by Fluorescence

The interaction between PUMA and MCL-1 was monitored by following the change in intrinsic tryptophan fluorescence of MCL-1 upon binding PUMA (Chapters 5.3 and 6.1.3). This typically resulted in a large amplitude change of up to ~15% of the starting signal. In contrast, upon mixing PUMA with A1, a relatively small signal change of only 0.6% was observed (Figure 8.4). A smaller signal results in a smaller signal to noise ratio. This can be overcome by averaging more traces, but this is more expensive, both in terms of time and protein. Additionally, mutations to the protein sequence are investigated during a φ-value analysis. Mutations can affect

the fluorescence intensity change, potentially making it unmeasurable, given the small signal for WT PUMA.

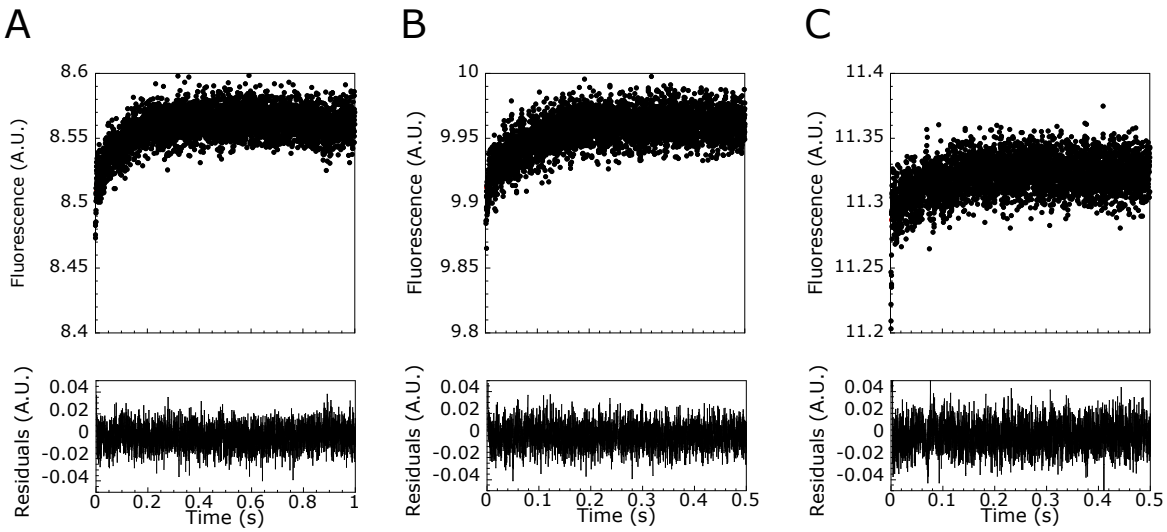


Figure 8.4 Fluorescence Intensity Changes Observed Upon Mixing PUMA with A1.
A small change in fluorescence intensity was observed upon rapidly mixing PUMA (0.25 μM) with various concentrations of A1. The reaction was performed under pseudo-first-order conditions with A1 at concentrations of (A) 2.5 μM, (B) 3.3 μM and (C) 4 μM. Changes in fluorescence with time were fit to a single exponential function. Data represent the mean of 18 – 27 individual traces.

Another way to improve signal to noise ratios is to use a different probe. Conjugation of a fluorescent dye provides a few distinct advantages over intrinsic tryptophan fluorescence. Firstly, both protein partners may have fluorescent tryptophan or tyrosine residues. This creates a high background fluorescence, especially when monitoring the reaction under pseudo-first-order conditions, as only a maximum of 10% of the excess protein is able to go into complex and contribute to the signal change. In contrast, a fluorescent dye can be specifically added to just one of the protein partners, minimising the background fluorescence. The reaction can then be set up with the unlabelled protein in excess, resulting in up to 100% of the dye-labelled peptide going into complex. The signal change is therefore maximised. Secondly, when using a fluorescent dye, both a change in fluorescence intensity and fluorescence anisotropy can be measured. Thus, if the

fluorescence intensity change is lost upon mutating the protein sequence, the reaction can instead be monitored by anisotropy.

Fluorescent dyes can have different properties and different prices. A lack of photo-bleaching over long-timescales was considered an important requirement, due to the relatively slow dissociation rate constants observed for BH3-only:BCL-2-like interactions (Chapter 6.1.3). Consequently, TAMRA was chosen as a starting candidate because of its reported photo-stability (Dempsey *et al.*, 2011) and competitive price.

Addition of TAMRA to the N-terminus of PUMA significantly improved the signal to noise ratio (Figure 8.5). As discussed in Chapter 6.1 and Figure 6.5, the presence of the dye did not alter the K_d of PUMA binding to MCL-1.

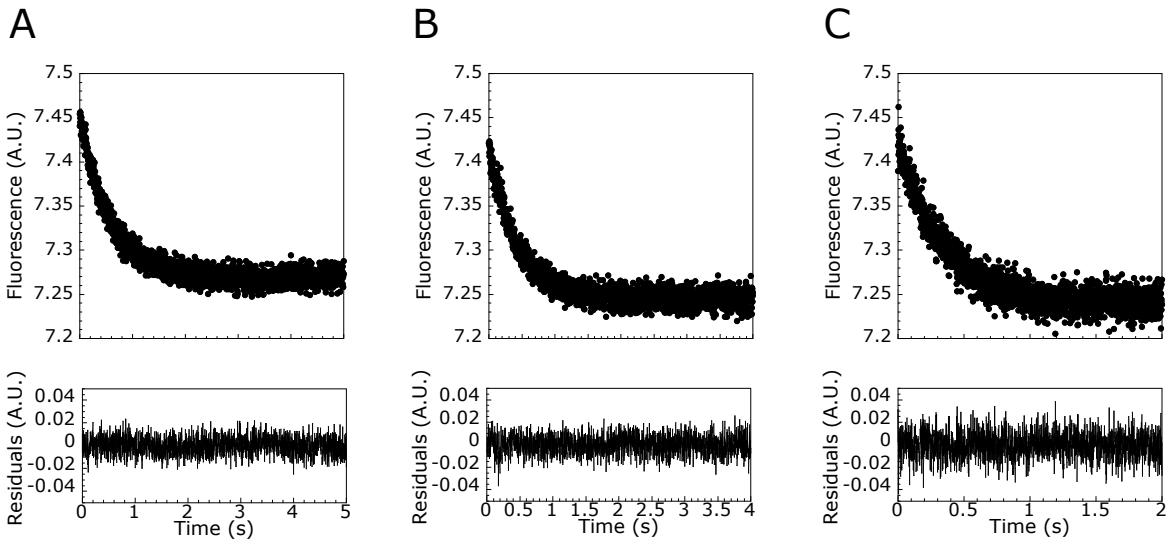


Figure 8.5 Fluorescence Intensity Changes Obtained Upon Mixing TAMRA-Labelled PUMA with A1.
A fluorescence intensity change of 2% of the starting signal was observed upon rapidly mixing PUMA (5 nM) with various concentrations of A1. The reaction was performed under pseudo-first-order conditions with A1 at concentrations of (A) 0.3 μM, (B) 0.4 μM and (C) 0.5 μM. Changes in fluorescence with time were fit to a single exponential function. Data represent the mean of 26 – 29 individual traces.

8.1.3 TAMRA Promotes PUMA Oligomersiation

Due to the improved signal change, TAMRA appeared to be a good probe of the reaction between PUMA and A1. However, TAMRA is hydrophobic. Oligomerisation of PUMA 34 (Figure 5.3) was reduced by mutating residue I137A in PUMA 35 (Figure 6.2). Unfortunately, addition of TAMRA at the N-terminus of PUMA 35 promoted the oligomerisation of the peptide. PUMA oligomerisation could be observed by several methods: concentration dependent CD spectra, the absorbance ratio at 555 nm compared to 523 nm and in kinetics (Figure 8.6). The potential to oligomerise was related to the amino acid sequence. Mutations that reduced the net charge of PUMA (i.e. negative to neutral mutations such as D147A) increased the oligomerisation propensity. This can be rationalised based on the reduced electrostatic repulsion upon decreasing the net charge. On the other hand, mutations that reduced helicity decreased the propensity to oligomerise (e.g. D147G displayed a reduced propensity to oligomersie compared to D147A).

In pseudo-first-order association kinetics experiments, biphasic kinetics were observed at certain concentrations of excess A1. The additional phase was relatively fast and had an opposite amplitude. The concentration of A1 at which the biphasic kinetics were observed was dependent on the oligomerisation propensity of the PUMA peptide. For example, D147A displayed biphasic kinetics above 1 μM A1, whereas others displayed single exponential kinetics at all concentrations. When two phases were detected, the observed rate started to reduce and the concentration dependence was no longer linear (Figure 8.7). These kinetic signatures indicated that another process was occurring at higher concentrations of A1. As the presence of these kinetic signatures was related to the oligomerisation propensity of PUMA, Occam's razor would suggest that the additional rate was due to the oligomer. One explanation could be that the oligomer is able to bind rapidly to A1 (giving the fast, additional rate), before dissociating and rebinding. The rate to reach the final equilibrium (a single PUMA chain bound to a single A1 protein) would be rate limited by the dissociation of the bound oligomer, producing the rollover in rate observed in Figure 8.7 panel C.

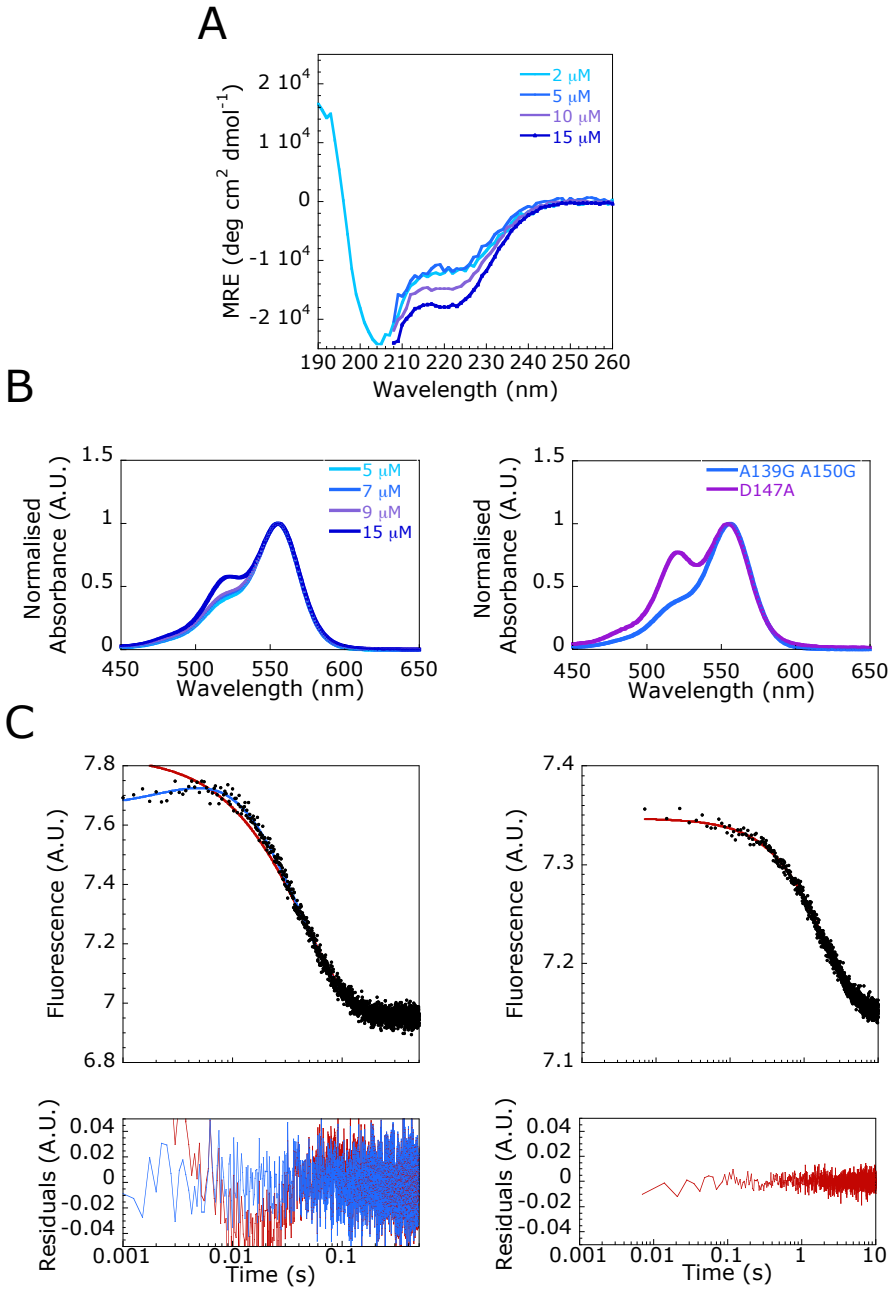


Figure 8.6 CD, Absorbance and Kinetic Signatures of PUMA Oligomerisation.

(A) A Concentration dependent CD spectrum is observed for PUMA WT, indicating oligomer formation in the low μM range. (B) Peak absorbance of TAMRA was observed at 555 nm. The ratio of absorbance at 555:523 nm is concentration dependent for PUMA WT (left). Non-oligomerising peptides, such as A139G A150G, display a concentration independent 555:523 ratio of ~2.5 (right). The difference in 555:523 ratio for D147A, which had the highest propensity to oligomerise, is indicated (right). (C) At high concentrations of excess A1 (5 μM; left), pseudo-first-order kinetic traces did not fit well to a single exponential (red), but were captured by a double exponential (blue). At low concentrations of A1 (0.1 μM; right), traces fit well to a single exponential.

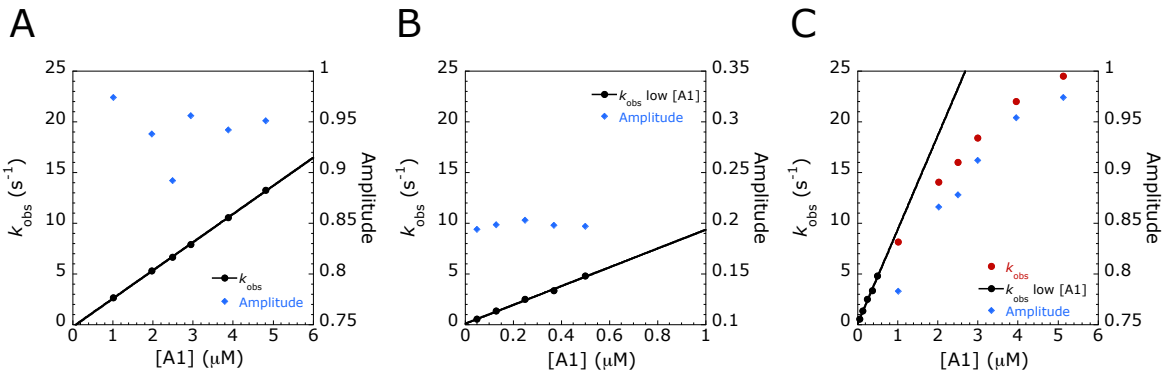


Figure 8.7 Rollover and Amplitude Changes Occur for Oligomerising Peptides.
TAMRA labelled PUMA peptides were rapidly mixed with an excess concentration of A1 and the rate to reach equilibrium (k_{obs}) was monitored. (A) A linear dependence of k_{obs} on the concentration of excess A1 was observed for A139G A150G, which displayed no propensity to oligomerise. (B) At low concentrations, D147A, the peptide with the highest oligomerisation propensity, displayed single exponential kinetics and a linear dependence of k_{obs} on the concentration of excess A1. (C) At higher, micro-molar concentrations of A1, biphasic kinetics were observed for D147A. In contrast to the low concentrations (black), k_{obs} was no longer linearly dependent on the A1 concentration (red). An increase in amplitude was also observed when linearity was lost. This change in amplitude was used as a diagnostic test and only traces with a single phase and similar amplitudes were used to determine k_{on}.

Monitoring the rate of monomeric PUMA association to A1 was the aim of the experiment and a requirement for φ-value analysis. For peptides which displayed a propensity to oligomerise, this could be achieved by performing the experiment at low concentrations of A1, where single exponential kinetics and a linear concentration dependence of the rate were observed. Any traces that showed biphasic kinetics, or an amplitude that was dependent on the concentration of A1, were excluded. There were a few exceptions to this amplitude exemption rule: some mutations reduced the affinity such that different amounts of PUMA went into complex with A1. When this occurred, the amplitude increased with the A1 concentration and the rates were included in the analysis (assuming single exponential kinetics were observed).

8.1.4 Characterisation of the Interaction Between WT PUMA and A1

Due to the dependence on protein concentrations, errors in protein concentration calculation or differences in protein purity could lead to discrepancies in determining k_{on}. This would be particularly problematic during a φ-value analysis, as mutants are typically characterised with different partner protein stocks. Consequently, WT was characterised with two independent stocks of A1. These stocks were produced at different times and had their concentrations independently measured. Both stocks showed good agreement, indicating that the data were reproducible with biological repeats (Figure 8.8).

The observed rate constants (k_{obs}) obtained from pseudo-first-order association experiments fit well to a straight-line (Figure 8.8). The gradient provides the k_{on}, while the k_{off} can be estimated from the intercept. As the k_{off} for the interaction was so slow ($5 \times 10^{-4} \pm 0.2 \times 10^{-4}$ s^{-1}), it cannot be accurately determined from the intercept. Nevertheless, the intercept (0.1 ± 0.1 s^{-1}) was not inconsistent with the k_{off} measured from out-competition experiments. A K_d of 85 ± 4 pM was determined from the ratio of k_{off} over k_{on}. Thus, PUMA binds slightly tighter to A1 than MCL-1 (300 pM K_i; Chapter 6.1.2.).

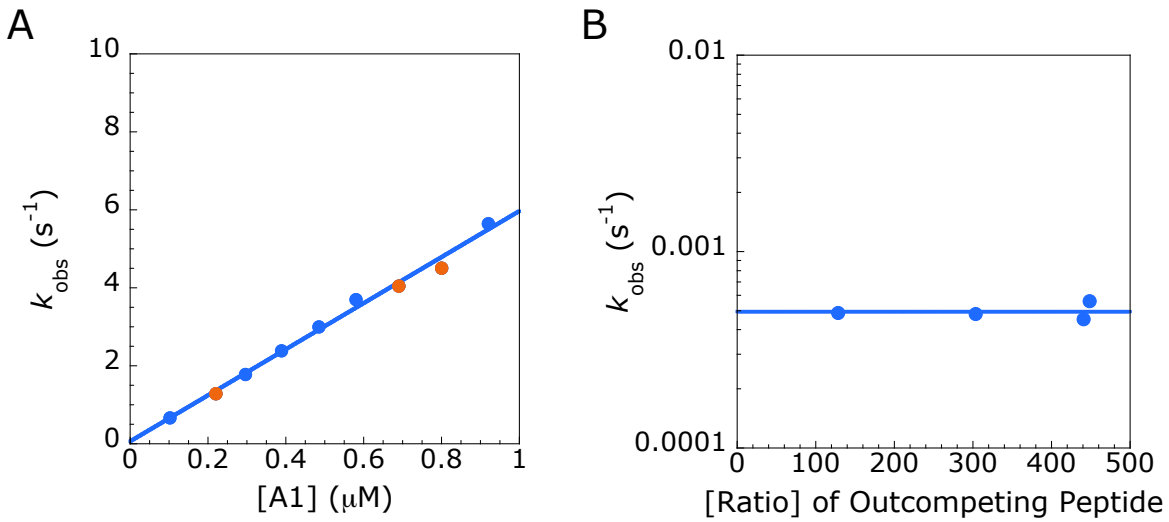


Figure 8.8 Reaction Kinetics for the Interaction Between PUMA and A1. (A) Observed rate constants (k_{obs}) were obtained from pseudo-first-order association kinetics experiments. Two independent stocks of A1 were used (blue and orange). An association rate constant of 5.9 ± 0.2 μM^{-1} s^{-1} was obtained from the gradient of the straight-line fit. (B) The change in fluorescence observed upon mixing a pre-formed complex of A1 and PUMA with an excess of unlabelled PUMA was fit to a single exponential to obtain k_{obs}. A dissociation rate constant of 5 x 10^{-4} ± 0.2 x 10^{-4} s^{-1} was determined from the mean of the concentration independent k_{obs}. Errors for k_{obs} are smaller than the data points. Errors in k_{on} represent the error of the fit, whereas the error in k_{off} represents the standard error of the mean (n=4).

8.1.5 Choice of Mutations

Fundamentally, φ-value analysis measures the contribution of Gibbs free energy for each residue at the transition state compared to the final, bound state. This is achieved through disrupting the interactions of the residue of interest through making a mutation. Mutations must therefore be sufficiently destabilising to allow differences between the WT and mutant to be discerned. Typically, the minimum change in $\Delta\Delta G$ required between the bound and free states is >0.6 $kcal.mol^{-1}$ (Fersht and Sato, 2004). With this in mind, and to match the mutations made in the PUMA:MCL-1 φ-value analysis, the hydrophobic residues in PUMA were mutated to alanine. These residues become buried into the hydrophobic grove of A1 upon binding. Shortening the side chain by mutation to alanine should remove some of the hydrophobic interactions with A1, resulting in a loss of binding affinity. Interactions with the side chains of residues i±4 can also be affected. These mutations therefore probe both intermolecular and intramolecular contacts.

The amount of helicity can be specifically probed by mutating residues that are solvent exposed and do not make intermolecular contacts in the bound structure. Charged residues may be present on the solvent exposed surface. Mutating these residues could lead to differences in k_{on} that are at least partially due to changes in long-range electrostatic steering. Consequently, analysis of helical content requires two mutations. First, the residue is mutated to an alanine. This sequence is used as a surrogate WT. Helicity is disrupted by mutating the same residue to glycine. Differences in free energy between the surrogate WT and the glycine mutant are then used to calculate φ. Six residues were investigated in this manner to probe the helix formation of PUMA in the transition state.

Another aspect to choosing mutations is that they should be relatively non-disruptive. If mutations are too drastic, they could cause significant destabilisation of the transition state. This could result in a second, distinct transition state becoming lower in energy than the mutated, original transition state. This alternative transition state may involve completely different interactions. A disruptive mutation would therefore be reporting on a different transition state to the one relevant to the WT sequence. To check that destabilising mutations do not disrupt the interactions of other residues in the transition state, a double mutant can be analysed. The double mutant should contain at least one mutation that alters the Gibbs free energy of the transition state. If the two mutations have been characterised individually, then the double mutant can be compared to the single mutant background. *i.e.* if a double mutant, A139G A150G, is produced, the φ-value for A150G in this context can be calculated from the difference in free energy between the double mutant and A139G. If the A139G mutation does not alter the interaction of A150G in the transition state, then the same φ-value should be obtained.

8.1.6 Effect of the Mutations on Residual Helicity

To ensure that the CD spectrum represented the monomeric peptide, scans were performed at low concentrations (≤1 μM) using a 1 cm cuvette. While the long path-length cuvette allowed these low concentrations to be measured, unfortunately it resulted in a large amount of background absorbance at low wavelengths. Consequently, to prevent the detector from exceeding its recommended voltage, scans were stopped at 200 nm. At least two concentrations

of each PUMA peptide were scanned to determine whether oligomers were present (Figure 8.9 and Figure 8.10). Only D147A showed a significant difference in CD spectrum between 0.5 and 1 μM (Figure 8.10). To check that the 0.5 μM spectrum was representative of the monomer, an additional scan was performed at 0.25 μM. This overlaid with the spectrum obtained at 0.5 μM, indicating that the helicity of the monomeric peptide could be estimated from the 0.5 μM spectrum.

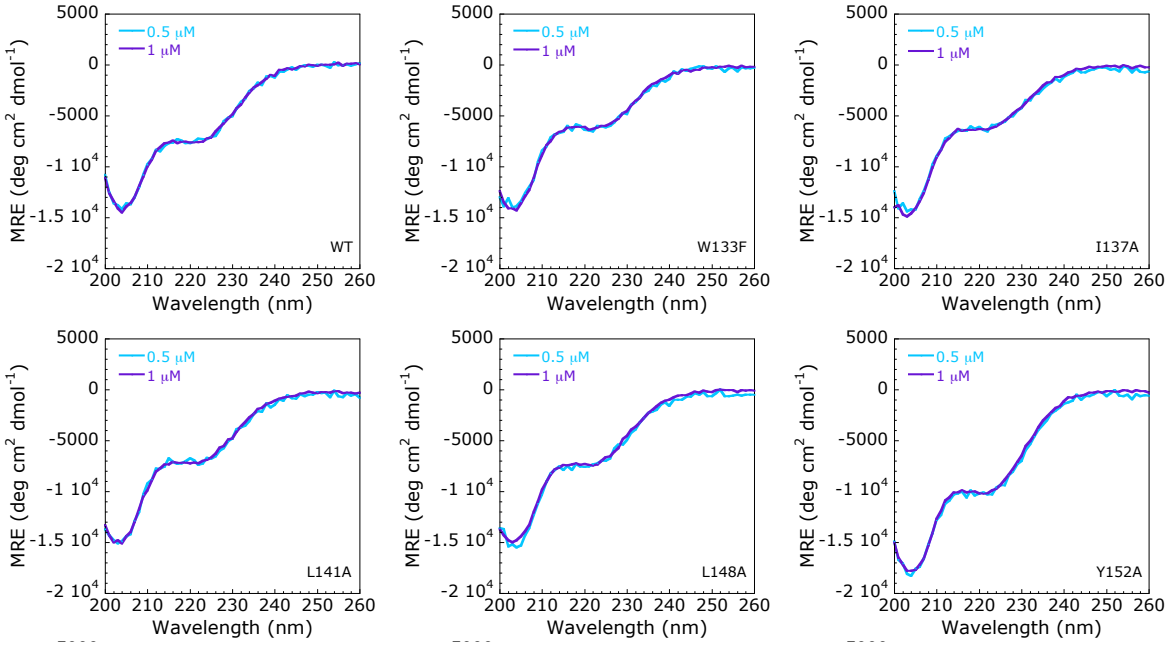


Figure 8.9 CD Concentration Dependence for Hydrophobic to Ala Mutations. Mutations to alanine were designed to disrupt the interaction of specific hydrophobic residues in PUMA with A1. CD was performed to determine the effect of the mutation on the residual helicity of PUMA. Peptides were scanned at two concentrations (0.5 and 1 μM) to determine the concentration dependence of the CD spectrum.

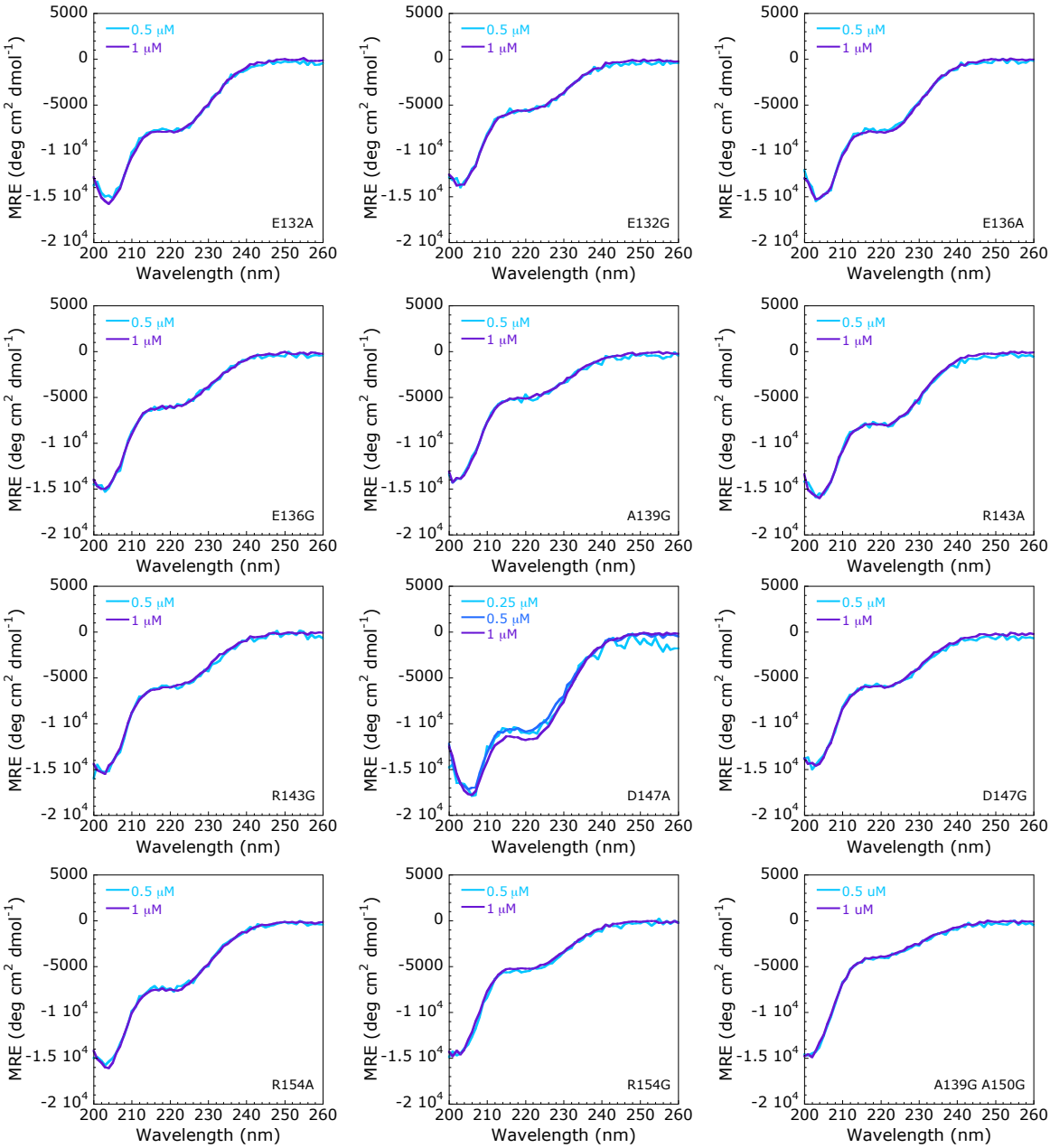


Figure 8.10 CD Concentration Dependence for Ala to Gly Mutants.
Ala to Gly mutations were performed to probe the helicity of PUMA within the PUMA:A1 transition state. At least two concentrations (0.5 and 1 μM) were scanned for each peptide to determine the concentration dependence of the CD spectrum. D147A demonstrated a different CD spectrum at 0.5 and 1 μM. Therefore, an additional scan at 0.25 μM was performed, which overlaid with the spectrum obtained at 0.5 μM.

Estimates of the residual helicity of each peptide were computed using the mean residual ellipticity value at 222 nm and the method of Muñoz and Serrano (1995). An increase in helicity of was observed for the R143A, D147A and Y152A mutations (Figure 8.11). All other mutants displayed similar or reduced helicity compared to WT. As expected, the residual helicity of glycine mutants was reduced compared to their alanine counterparts. Helicity estimates for the TAMRA labelled PUMA peptides investigated in this work generally matched those obtained for the unlabelled versions. Addition of the dye, therefore, did not appear to significantly affect the residual structural of PUMA.

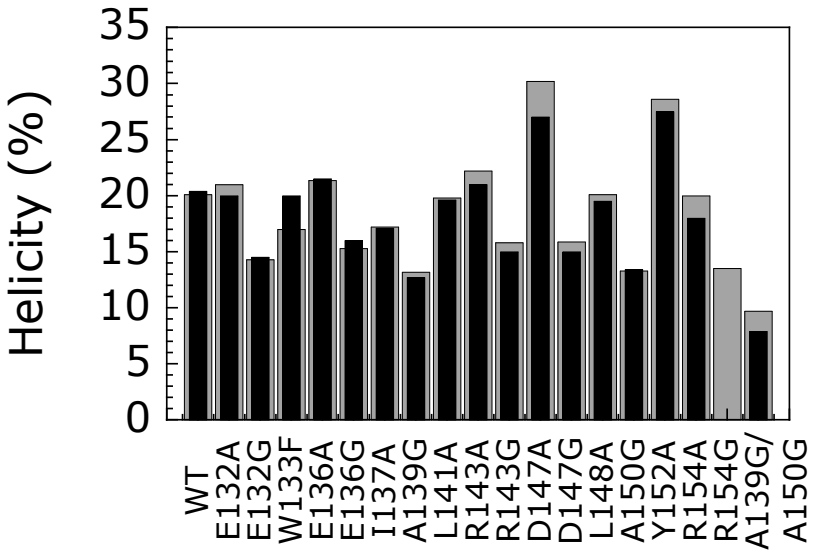


Figure 8.11 Helicity Estimates for PUMA Peptides.
Estimates of the helicity were obtained from the MRE at 222 nm. As expected, glycine mutants typically reduced the residual helicity compared to the equivalent alanine mutant. Grey bars represent the helicity obtained for the TAMRA labelled peptides investigated in this work. Black bars represent the helicity obtained by Rogers *et al.* (2014) for the corresponding unlabelled peptides. No previous estimate was obtained for R154G.

8.1.7 φ-Values

To determine the Gibbs free energy changes upon mutation, a kinetic analysis of the interaction between the PUMA peptides and A1 was undertaken. PUMA and A1 were rapidly mixed using stopped flow and the change in fluorescence intensity or anisotropy with time was monitored. The reactions were set up under pseudo-first-order conditions, with A1 in excess. Signal changes were fit to a single exponential function to obtain the observed rate constant (k_{obs}). To maintain a linear dependence of k_{obs} on the concentration of A1, kinetics for WT, E136A, D147A, D147G and L148A were performed at low concentrations. Association rate constants were obtained from the gradient of the straight-line fit of k_{obs} against the concentration of A1 (Figure 8.12 and Figure 8.13). The concentration range used to determine the k_{on} affected the error, but did not alter the estimate. For example, a k_{on} of 3.1 ± 0.3 $\mu M^{-1}s^{-1}$ was obtained for A139G when a concentration range of 0.1 - 1 μM was used. When the range of concentrations was extended to 10 μM, the value of k_{on} was the same within error (3.26 ± 0.05 $\mu M^{-1}s^{-1}$).

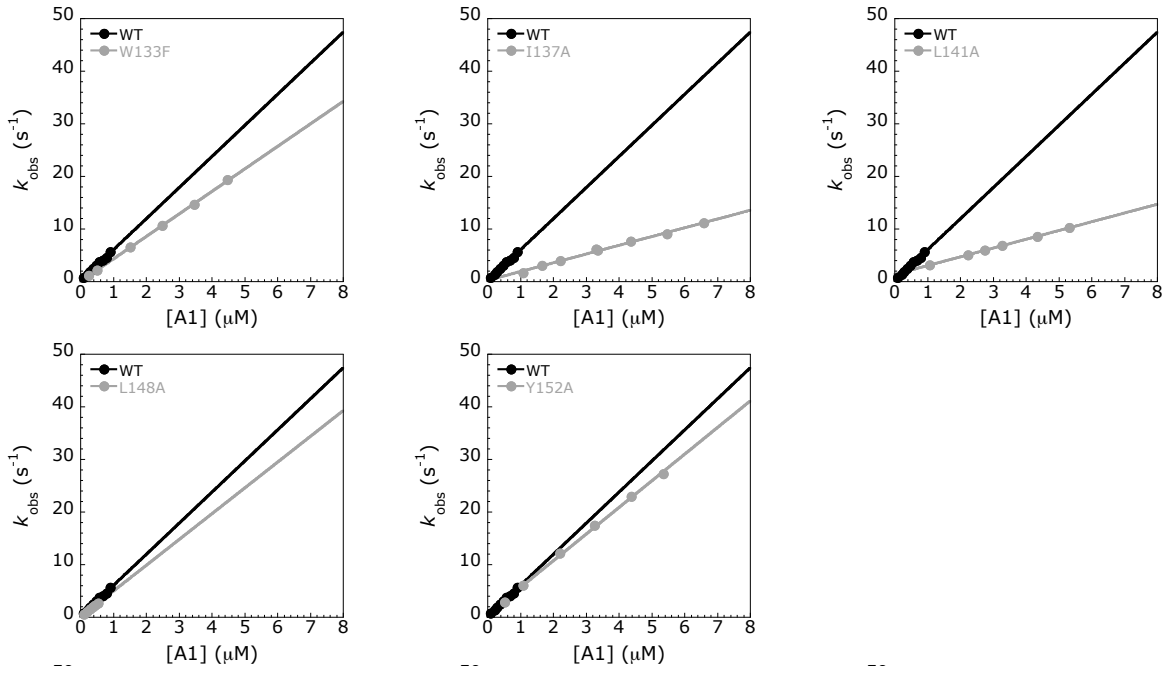


Figure 8.12 Association Kinetics for PUMA Hydrophobic to Alanine Mutants. PUMA was rapidly mixed with various concentrations of excess A1. Fluorescence intensity changes were fit to a single exponential to obtain k_{obs}. The gradient of the straight-line fit of k_{obs} against the A1 concentration provides the association rate constant (k_{on}). WT is shown in black, mutants are in grey.

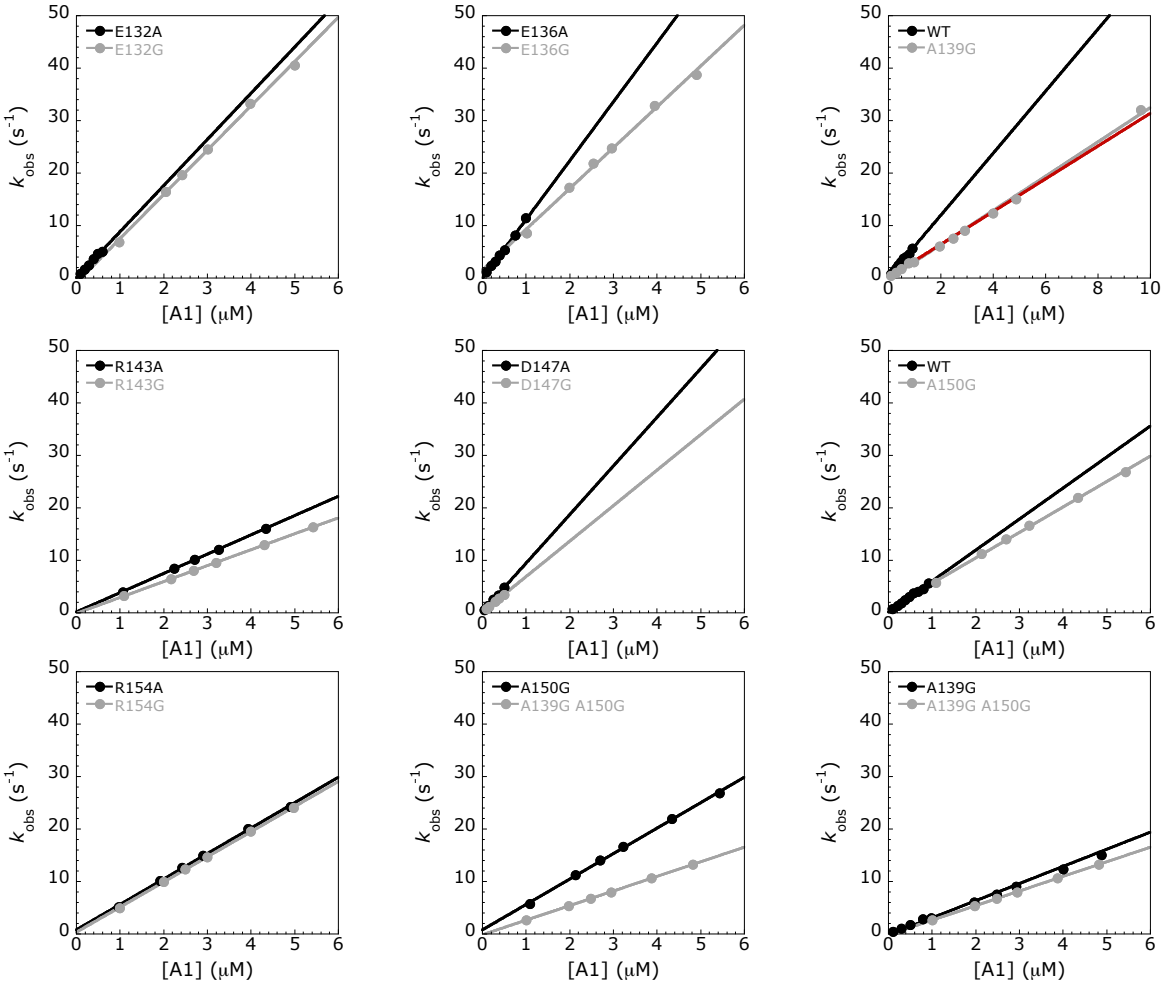


Figure 8.13 Association Kinetics for PUMA Alanine to Glycine Mutants.
PUMA was rapidly mixed with various concentrations of excess A1. Fluorescence intensity changes were fit to a single exponential to obtain k_{obs}. The gradient of the straight-line fit of k_{obs} against the A1 concentration provides the association rate constant (k_{on}). WT, or the surrogate alanine WT are shown in black, mutants are in grey. A wider concentration range was examined for A139G. The same k_{on} was obtained when a concentration range of 0.1 – 1 μM (red line) or 0.1 – 10 μM (grey line) was used.

Dissociation kinetics were monitored by mixing a pre-formed complex of dye-labelled PUMA peptide and A1 with an excess of unlabelled WT PUMA. The change in fluorescence with time was fit to a single exponential to obtain k_{obs}. When the competing peptide concentration was high enough, the k_{obs} became independent of the competing peptide concentration (Figure 8.14 and Figure 8.15). The mean average of these concentration independent k_{obs} provides the dissociation rate constant (k_{off}). In keeping with the interaction between PUMA and MCL-1 (Rogers *et al*., 2014a), mutation of L141A provided the largest change in k_{off}. Combined with a reduced k_{on}, this mutation shifted the affinity by almost 4-orders of magnitude. For comparison, all other mutations altered the affinity by <3-orders of magnitude.

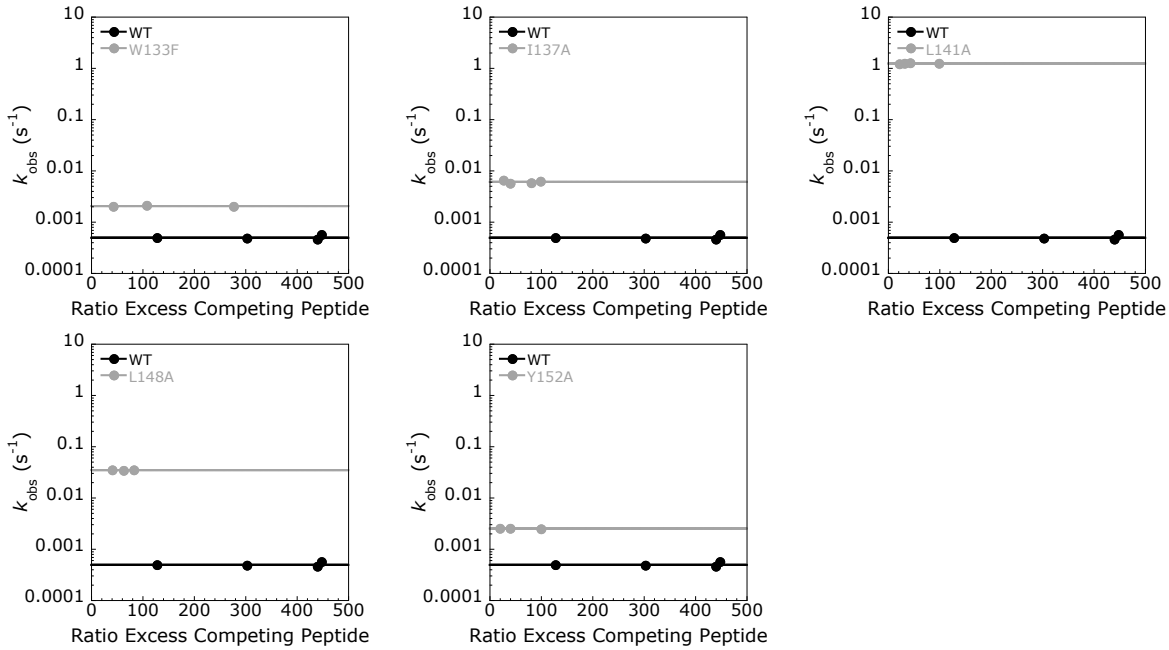


Figure 8.14 Dissociation Kinetics for PUMA Hydrophobic to Alanine Mutants.

A pre-formed complex of dye-labelled PUMA and A1 was mixed with various excess concentrations of unlabelled WT PUMA. Fluorescence intensity changes were fit to a single exponential to obtain k_{obs}. Dissociation rate constants were obtained from the mean of the concentration independent k_{obs}.

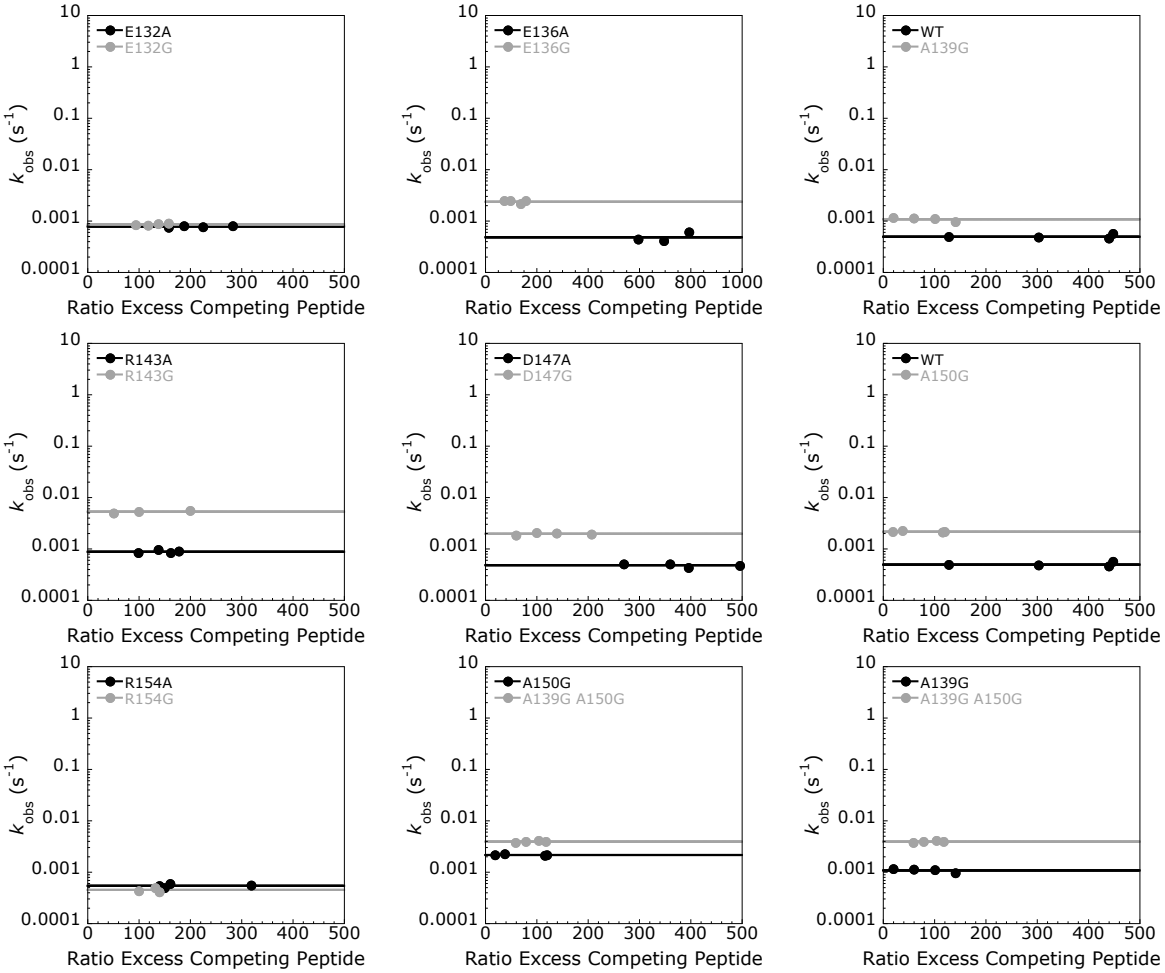


Figure 8.15 Dissociation Kinetics for PUMA Alanine to Glycine Mutants.
A pre-formed complex of dye-labelled PUMA and A1 was mixed with various excess concentrations of unlabelled WT PUMA. Fluorescence intensity changes were fit to a single exponential to obtain k_{obs}. Dissociation rate constants were obtained from the mean of the concentration independent k_{obs}.

φ-values can be calculated from the kinetic data using Equation 8.1:

$$\varphi = \frac{\ln\left(\frac{k_{on}^{WT}}{k_{on}^{Mut}}\right)}{\ln\left(\frac{K_d^{Mut}}{K_d^{WT}}\right)} \quad (8.1)$$

The K_d in the denominator of Equation 8.1 can be calculated by taking the ratio of the k_{off} over the k_{on} (Equation 2.21). This technique of calculating K_d holds as long as the reaction is 2-state. To check this assumption, equilibrium binding curves for the two most destabilising mutants, L141A and L148A were performed. These mutations were chosen as the affinity was shifted to a range that was amenable to the assay. Higher affinity interactions would require lower protein concentrations, shifting the assay into a regime where proteins sticking to plastic and glassware starts to become an issue. Equilibrium dissociation constants of 5.65 ± 0.05 nM for L148A and 710 ± 20 nM for L141A were obtained from the mean of two repeats (Figure 8.16). These compared well with the values of 7.10 ± 0.09 nM (L148A) and 735 ± 7 nM (L141A) calculated from kinetics. φ-values for all mutants were therefore calculated from the kinetic data (Table 8.1).

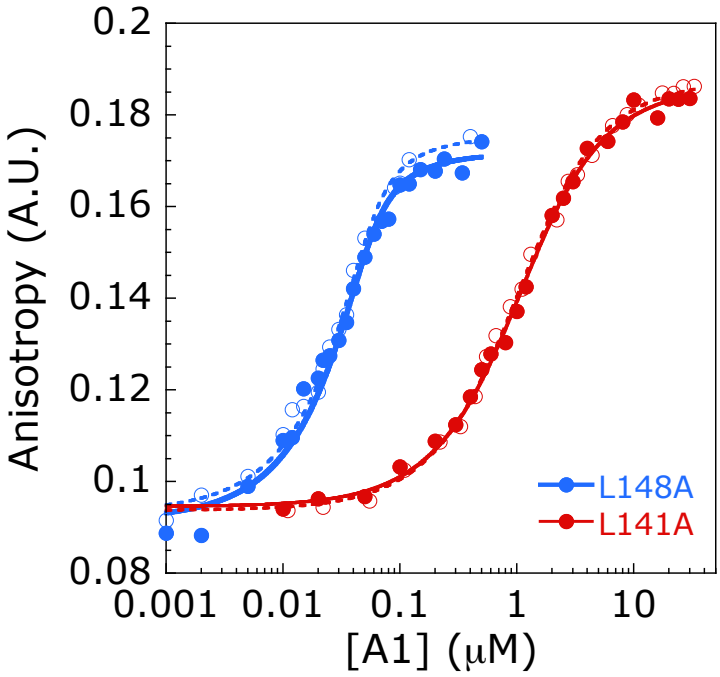


Figure 8.16 Equilibrium Binding Curves for PUMA L141A and L148A. Various concentrations of A1 were incubated with 640 nM of PUMA L141A and 50 nM of L148A. Two repeats are shown for each binding curve. To aid the eye, one repeat is shown with open circles and a dashed line, the other with closed circles and an unbroken line. Data were fit to Equation 3.15 to obtain the K_d.

Table 8.1 Kinetic Data and φ-Values for PUMA Binding to A1.

PUMA	k_{on} ($\mu M^{-1}s^{-1}$)	k_{on} Error	k_{off} (s^{-1}) x10^{-3}	Error (SEM) ±	n	K_d (k_{off}/k_{on}) (nM)	Error ±	ΔΔG (kcal. mol^{-1})	Error ±	φ	Error ±
WT	5.9	0.2	0.5	0.02	4	0.085	0.004	-	-	-	-
E132A	8.8	0.5	0.77	0.01	4	0.088	0.005	0.02	0.05	-	-
E132G	8.4	0.2	0.86	0.01	4	0.102	0.003	0.09	0.03	0.30	0.02
W133F	4.27	0.04	2.03	0.03	3	0.475	0.008	1.02	0.03	0.19	0.02
E136A	10.4	0.2	0.48	0.04	4	0.046	0.004	-0.36	0.06	-	-
E136G	8.2	0.2	2.38	0.09	4	0.29	0.01	1.09	0.04	0.13	0.01
I137A	1.64	0.05	6	0.2	4	3.7	0.2	2.23	0.04	0.34	0.01
A139G	3.26	0.05	1.08	0.04	4	0.33	0.01	0.81	0.04	0.44	0.02
L141A	1.67	0.01	1227	8	4	735	7	5.37	0.03	0.139	0.003
R143A	3.67	0.04	0.88	0.03	4	0.240	0.008	0.62	0.04	-	-
R143G	3.03	0.01	5.2	0.2	3	1.72	0.07	1.17	0.04	0.097	0.002
D147A	9.4	0.3	0.48	0.02	4	0.051	0.003	-0.92	0.04	-	-
D147G	6.76	0.08	1.95	0.05	4	0.288	0.008	1.03	0.04	0.2	0.1
L148A	4.90	0.05	34.8	0.3	3	7.10	0.09	2.62	0.03	0.042	0.008
A150G	4.84	0.08	2.15	0.02	4	0.444	0.008	0.98	0.03	0.12	0.02
Y152A	5.1	0.1	2.51	0.03	3	0.49	0.01	1.04	0.03	0.08	0.02
R154A	4.86	0.08	0.54	0.02	4	0.111	0.005	0.16	0.04	-	-
R154G	4.79	0.03	0.44	0.03	3	0.092	0.006	-0.11	0.05	-0.076	0.002
A139G (A150G)	2.78	0.02	3.9	0.1	4	1.40	0.04	0.68	0.03	0.48	0.01
A150G (A139G)	2.78	0.02	3.9	0.1	4	1.40	0.04	0.82	0.03	0.11	0.01

E132G was the only mutation that did not satisfy the >0.6 kcal.mol^{-1} change in affinity required to calculate φ-values. All other mutations generated usable φ-values. Looking at Table 8.1, it is clear that the transition state of PUMA:A1 is characterised by generally low phi-values, with few interactions between PUMA and A1 formed at this point of highest energy. The higher φ-values that were observed were clustered towards the N-terminus, indicating that this region is important for guiding the reaction. At the C-terminus, the pattern of φ-values was essentially 0, suggesting that this region is not forming native-like interactions at the transition state.

Double mutants were produced as a control to ensure that the mutations were not causing large scale disruptions at the transition state. Both double mutants gave essentially identical φ-values (0.44 ± 0.02 *vs* 0.48 ± 0.01 for A139G; and 0.12 ± 0.02 *vs* 0.11 ± 0.01 for A150G), indicating that destabilisation of the transition state by one mutation was not affecting the interactions of other residues.

8.2 Discussion

Broadly speaking, a similar pattern of φ-values was observed for PUMA binding to both A1 and MCL-1. Both transition states were characterised with a lack of high φ-values, indicating that few native-like interactions were present. This is consistent with the early, largely unstructured transition states observed in the majority of other studies of IDPs that undergo coupled folding and binding (Dogan *et al*., 2013; Haq *et al*., 2012; Rogers *et al*., 2014a; Toto and Gianni, 2016).

When binding to either A1 or MCL-1, PUMA displayed higher φ-values towards the N-terminus. Strikingly, the similarity extends beyond just high or low categories. Virtually all the calculated φ-values for PUMA were identical, irrespective of the binding partner (Figure 8.17). The consistency of these φ-values indicates that the IDP may encode which interactions form in the transition state. If this was the case, a different BH3-only peptide, such as BID, should demonstrate a distinct pattern of φ-values. Furthermore, if the partner protein has no influence on the transition state interactions, the distinct φ-values should be reproduced when binding to either MCL-1 or A1. A φ-value analysis of BID was therefore performed to test this hypothesis.

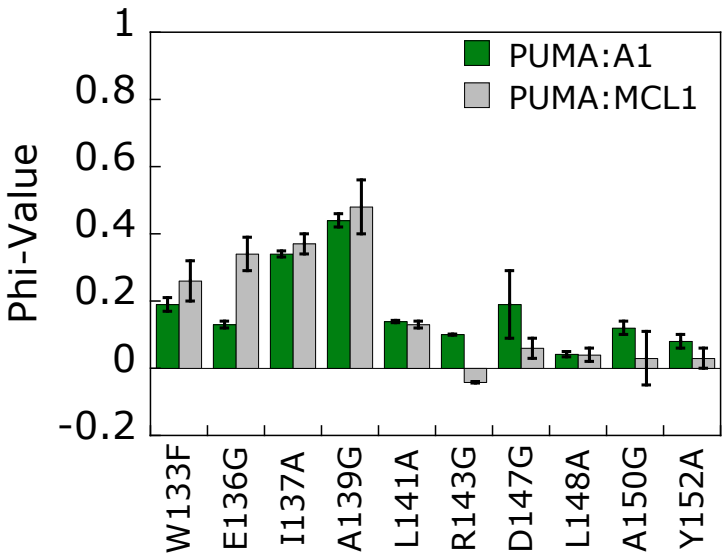


Figure 8.17 Comparison of PUMA Phi-Values When Binding to A1 or MCL-1. Data for PUMA:MCL-1 were taken from Rogers *et al.* (2014). A φ-value of 0 indicates that no native-like structure is formed in the transition state, whereas a value of 1 indicates that all of the native interactions are formed in the transition state. Intermediate φ-values indicate that the residue is either making some, but not all, of its native-like interactions in the transition state. Or the residue is making all of its native interactions, but with a lower energy than in the bound state.

As PUMA required the addition of a fluorescent dye to follow the reaction kinetics with A1, the BH3-only peptide, BID was also purchased with an N-terminal TAMRA dye. PUMA and BID were therefore more comparable and the dye provided the additional safety net of being able to monitor either fluorescence intensity or anisotropy. Fortunately, BID peptides did not display any propensity to oligomerise in the concentration ranges examined. After performing the initial characterisation, the interactions between BID and MCL-1 and BID and A1 were investigated by Quenton Bubb and Carolina Mendonça, respectively.

As for PUMA, mutations were chosen on each turn of the helix. Both hydrophobic residues that become buried into the grove of A1 and MCL-1, and surface, solvent exposed residues were investigated. Only five residues provided a $\Delta\Delta G$ above the 0.6 kcal.mol^{-1} typically required to calculate φ-values. All of these mutations were to hydrophobic residues that become buried upon binding. Four out of the five residues were in equivalent positions to the amino acids investigated in PUMA

(Figure 8.18). The additional BID mutant (I83A) was on the same helical turn as PUMA W133, but was in the equivalent position of PUMA residue 134. BID I82A was investigated, however, the $\Delta\Delta G$ was 0.4 kcal.mol^{-1} for BID binding to MCL-1 and 0.5 kcal.mol^{-1} for BID binding to A1.

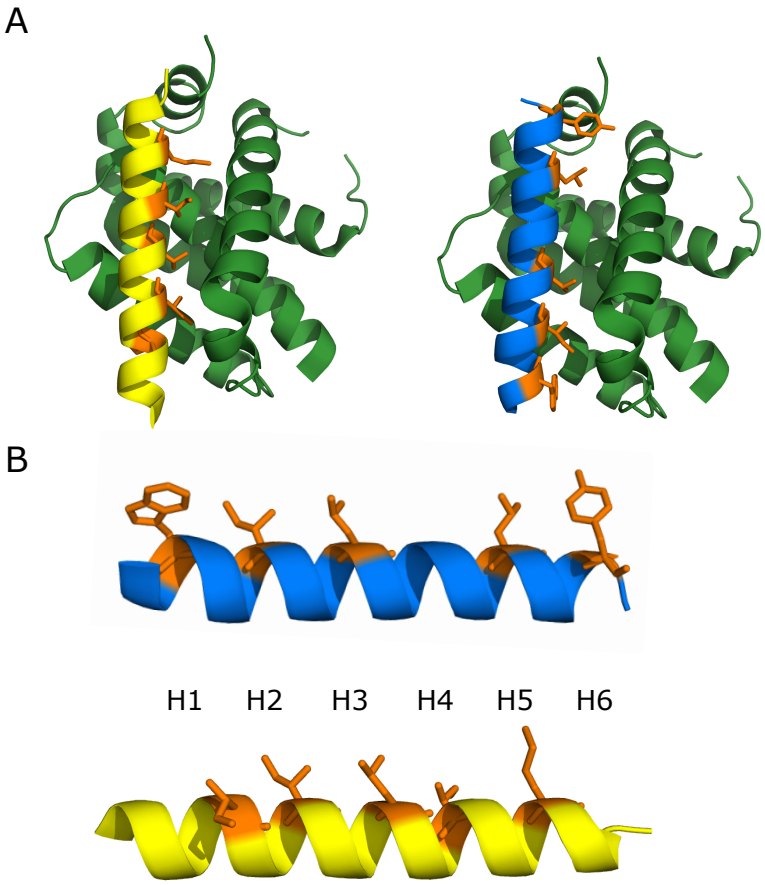


Figure 8.18 Position of Mutated Residues in PUMA and BID.
(A) Structures of BID (yellow) and PUMA (blue) bound to A1 (green). (B) Each mutant is on a separate turn of the helix. For comparison, the hydrophobic mutants (orange sticks) are referred to relative to their position in the helix. *i.e.* the N-terminal residues W133 (PUMA) and I83 (BID) are indicated at position H1. Note that position H4 in PUMA is an alanine and was therefore not investigated.

As predicted, BID displayed a different φ-value pattern to PUMA (Figure 8.19 A). Whilst PUMA displayed the highest proportion of native-like interactions at the N-terminus, BID had a φ-value of 0 for the most N-terminal residue. The φ-values for BID were relatively low throughout the entire helix; however, on average, the

central portion displayed slightly higher values. This indicates that the central region of BID is important in guiding the reaction to the transition state.

When binding to either MCL-1 or A1, the pattern of BID φ-values were highly similar. In contrast, when the partner protein is kept constant and the IDP is changed (Figure 8.19 B), it is clear that there is variation in the amount of native interactions each residue is making. The results obtained for BID are therefore consistent with the hypothesis that was derived from the PUMA MCL-1/A1 φ-value comparison.

8.3 Conclusion

For a unimolecular protein folding reaction, Anfinsen showed that all of the information required for folding was contained within the protein sequence (Anfinsen, 1973). At the start of this project, the aim was to determine which component of a coupled folding and binding reaction encodes the transition state information. The comparative φ-value analyses described here are consistent with the IDP containing the folding pathway instructions. Thus, analogous to Anfinsen's observation, it is the sequence of the protein that folds which encodes its transition state.

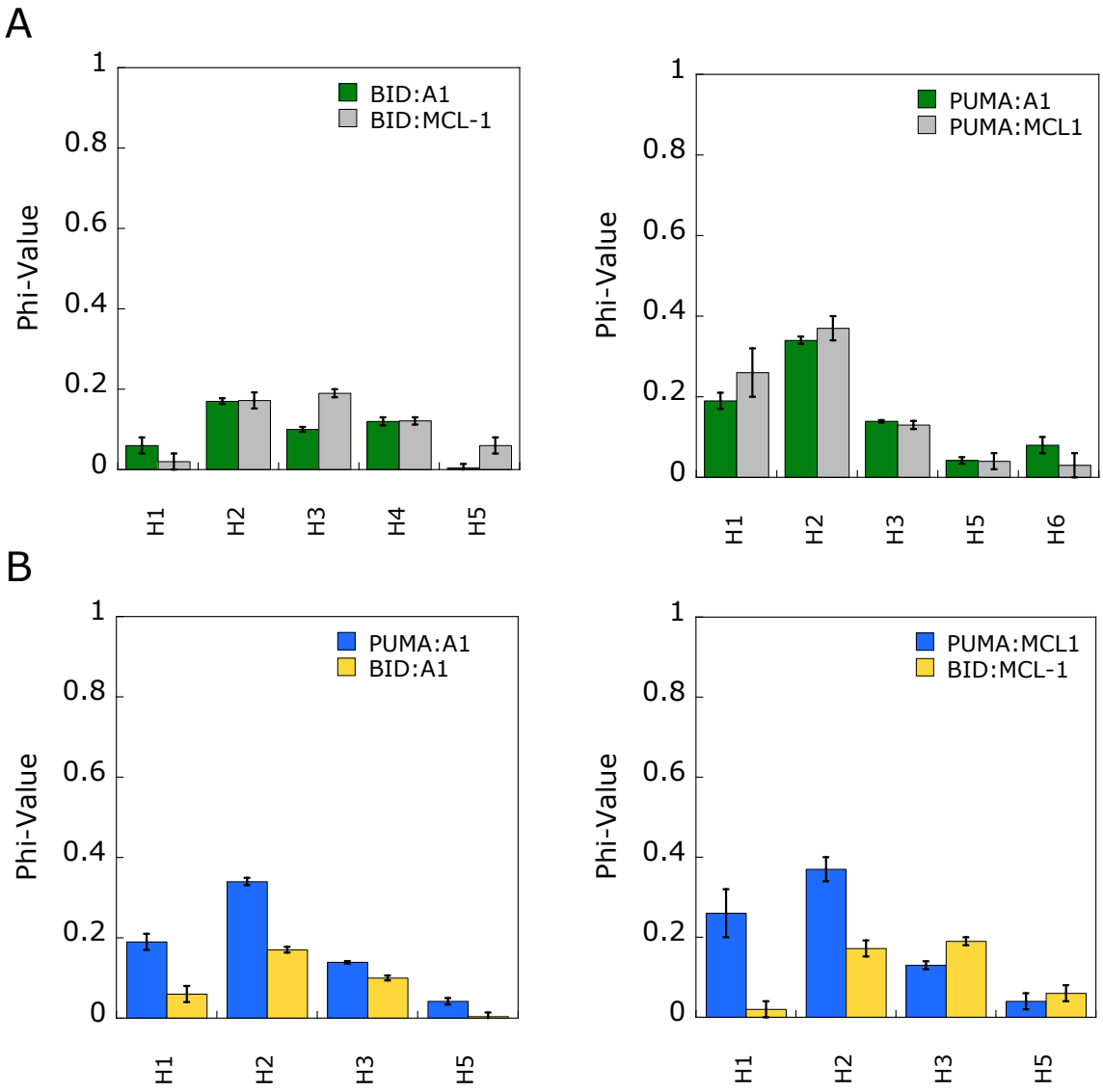


Figure 8.19 BID Displays a φ-value Pattern that is Distinct from PUMA. (A) A similar pattern of φ-values was obtained when either the IDP, BID (left) or PUMA (right), is kept constant and the binding partner is changed. (B) In contrast, when the partner protein is kept constant and the IDP is changed, the φ-values do not match and fall outside of error. Labels on the x-axis relate to the position of the residue, as indicated in Figure 8.18.

Chapter 9

Conclusions, Progress and Future Directions

Since this work began in 2014, the field of intrinsically disordered proteins has gathered momentum, expanded and moved forward impressively. A few of the developments are described below. The findings of this thesis are then summarised and potential future work is discussed.

9.1 Structural Ensembles

Intrinsically disordered proteins lack a well-defined three-dimensional structure and exist as an ensemble of conformations. As discussed in this thesis, a subset of IDPs can fold upon binding to a partner macromolecule. At the time of starting this work, the link between a structural ensemble and a well-defined folded state raised an interesting question: do IDPs bind in an already folded state, following a conformational selection mechanism, or is the folding induced by binding? If the first mechanism was true, then the folded state should be at least partially populated within the structural ensemble of the unbound IDP. Traditional techniques to determine protein structures required averaging over many molecules. This creates something of an issue if each molecule is transiently exploring different structural states (Mittag and Forman-Kay, 2007). To overcome these problems, new experimental methods were developed (Aznauryan *et al.*, 2016; Ozenne *et al.*, 2012), allowing structural populations in IDP ensembles to be described. Just as the structures of folded proteins are deposited in the Protein Data Bank, a new database was developed for IDP structural ensembles (Varadi *et al.*, 2014). So far, 24 entries have been deposited in the protein ensemble data base (PED).

9.2 Electrostatics

Through electrostatic steering, the association of folded proteins can be accelerated by 4-orders of magnitude (Schreiber and Fersht, 1996). Given that IDPs typically display an excess of charged residues (Romero *et al.*, 2001; Theillet *et al.*, 2014; Uversky *et al.*, 2000), it would be easy to assume that long-range electrostatic

interactions may have an even greater influence on the association of complexes that involve a disordered component. However, the opposite has been shown to be the case (Papadakos *et al.*, 2015). Association of the folded proteins - colicin E3 rRNase domain and Im3 - is electrostatically enhanced by 3 orders of magnitude (Walker *et al.*, 2003). Yet, when the colicin E3 rRNase domain was unfolded by mutating a core tyrosine residue to an alanine, the electrostatically driven association was largely lost (Papadakos *et al.*, 2015). Only a single order of magnitude enhancement remained. This is quite a staggering change, considering that the charge composition of the WT and tyrosine to alanine mutant were identical. One explanation could be that, to achieve the most efficient rate enhancement, defined charged patches are required. While folded proteins can achieve a relatively fixed patch, through bringing charged residues together during folding, the disordered, dynamic nature of IDPs makes this more difficult. Steering between complementary charge patches may therefore be lost in IDPs, explaining the reduced electrostatic enhancement. While this is an intriguing hypothesis, further analyses of electrostatic contributions to IDP association rates are required to support it.

9.3 Sensitivity to Denaturants

The contribution of electrostatic steering to association rate constants can be assessed by measuring k_{on} at different salt concentrations. Typically, the salt of choice is either NaCl or KCl. In a recent study, the effects of using other salts to screen long-range charge interactions were investigated (Wicky *et al.*, 2017). As well as altering k_{on} through modifying electrostatic steering, the salt manipulated the residual structure of the IDP. The amount of structural change was dependent on the ion-type and followed the Hofmeister series (Hofmeister, 1888). In contrast to folded proteins, where the influence of Hofmeister salts on structure is typically observed in the molar-range, ion-type specific effects are observed at concentrations as low as 10 mM. The lack of structural stability of IDPs therefore may make them more susceptible to small changes in environment. Given the variation in cell compartment ion composition and the involvement of ions in signalling (Clapham, 2007), this sensitivity of IDPs could be of functional relevance.

9.4 Allostery and Cooperativity

Just as IDPs can be induced to fold by partner macromolecules, neighbouring domains can influence the folding of disordered proteins. For example, SasG (*Staphylococcus aureus* surface protein G) is composed of two domains, which are both predicted to be intrinsically disordered. In isolation, the G5 domain is folded, whereas the E domain is disordered (Gruszka *et al.*, 2012). The E domain remains unfolded when it is present at the C-terminus of G5. However, when the E domain is tandemly expressed at the N-terminus of G5, interactions between the two domains induce E to fold. Furthermore, when G5 is destabilised and unfolded by mutation, the presence of E at the N-terminus stabilises both domains and results in folding (Gruszka *et al.*, 2016).

SasG is expressed on the cell surface of *Staphylococcus aureus*, where it forms a rod-like structure. Interestingly, when force is applied to the multidomain protein, the E-domains unfold at lower forces than the G5-domains (Gruszka *et al.*, 2015). Thus, the destabilised, disordered nature of the E-domain allows the multidomain protein to transition between a rigid rod-like structure and a flexible state under mechanical force. This could be a crucial functional advantage to SasG, which is exposed to the extracellular environment.

IDPs are over represented in processes such as cell signalling (Ward *et al.*, 2004). To prevent disease, signalling processes must be tightly regulated. One example of this is in the hypoxic response, which is mediated by the interaction between HIF-1α and the TAZ1 domain of CBP. The protein complex formed between HIF-1α and TAZ1 regulates the transcription of various genes, including CITED2. CITED2 can also bind to TAZ1, competing for binding with HIF-1α, and producing a negative feedback loop. Although both CITED2 and HIF-1α have almost identical affinities for TAZ1, sub-stoichiometric concentrations of CITED2 are able to outcompete HIF-1α. This apparent enhanced affinity is due to a secondary interaction site for CITED2. Upon binding this allosteric site, CITED2 causes a change in HIF-1α k_{off}, promoting HIF-1α dissociation (Berlow *et al.*, 2017). Allosteric effects have also been observed in other IDP systems (Shammas *et al.*, 2014), and could comprise an important regulatory feature of signalling complexes.

9.5 Phase Transitions

Compartmentalisation allows cells to separate and optimise specific cellular processes. For example, eukaryotic cells use membrane enclosed nuclei to localise the machinery required for DNA and RNA synthesis. In recent years, intrinsically disordered proteins have been shown to form liquid-like membrane-less organelles (Mitrea and Kriwacki, 2016). Under certain conditions, these IDPs can interact and form a phase-separated compartment through de-mixing with the surrounding solvent (e.g. the cell cytoplasm) (Brangwynne *et al.*, 2015).

One example of a phase separating protein is Ddx4, which contains a folded DEAD-box helicase domain surrounded by intrinsically disordered regions (Nott *et al.*, 2015). Cation-pi interactions, between arginine and phenylalanine residues in the disordered N-terminus of Ddx4 molecules, are crucial to overcome the entropic cost of de-mixing. Phase separation of Ddx4 is therefore dependent both on the protein concentration and the ionic strength of the solvent. Reversible de-mixing can be achieved by changing either of these factors, or the temperature. Functionally, the Ddx4 liquid droplets can melt double stranded DNA, stabilise single stranded nucleic acid structures and selectively absorb or exclude proteins (Nott *et al.*, 2016).

The ability to reversibly form functional membrane-less liquid droplets in response to specific stimuli, provides the cell with a distinct advantage. Changes in environmental conditions can result in new cellular compartments without the need for generating membrane bound organelles. Although the field is relatively new, phase transitions represents an exciting new area of intrinsically disordered protein research.

9.6 Conclusions of this thesis

9.6.1 Changes in Binding Affinity are Predominantly Due to Alterations in k_{off}

The correlation between increases in residual 'bound-like' structure and binding affinity generated interest in studying the unbound IDP. If coupled folding and binding reactions were to be fully understood, characterising the ensemble of unbound IDP conformations was thought to be important. However, if the presence of a 'bound-like' conformation was crucial for the reaction, then altering the residual structure should modulate the affinity by changing k_{on}. In this thesis, changes in residual structure had a greater impact on the k_{off}, rather than the k_{on}. This was irrespective of the method used to alter the conformation (point mutation, shuffling or swapping residues). Hence, it was the effect on the bound-state, not the free-state, that was responsible for changes in affinity.

Whilst it is easy to imagine that changes in the reactant can alter k_{on}, one of the main messages from this thesis is that the backwards reaction has to be considered. Even if no observable change in structure occurs in the bound complex, changes in residual structure can alter the entropic cost of folding. For IDPs with relatively early transition states, this will lead to the effect being predominantly on k_{off}. Consequently, it is important to consider the effects on the bound complex, rather than just interpreting changes in the unbound IDP.

9.6.1.1 Evidence for a Dock and Coalesce Binding Mechanism

One suggested mechanism for the folding and binding of IDPs is the dock and coalesce mechanism (Ou et al., 2017; Zhou et al., 2012). In this scheme, a segment of the IDP 'docks' with the partner macromolecule. The rest of the IDP then folds and 'coalesces' to the final bound structure. The rate limiting step for the formation of the complex is the IDP segment and partner macromolecule colliding and interacting. Conversely, for the complex to dissociate, the intra and intermolecular contacts formed during the coalescing stage need to be broken and the IDP needs to return to its relatively unfolded state. All of the systems studied in this thesis are therefore consistent with this binding mechanism, as any change in residual structure influenced k_{off} (the potential coalescing step) more than on k_{on}

(the potential docking step). It remains to be seen whether this will prove to be a general mechanism for IDP coupled folding and binding.

9.6.2 Non-contacting Residues Alter Binding Affinity

The prevalence of IDPs in cell signalling makes them attractive targets for drug development. Typically, binding affinity is enhanced through engineering the binding interface. Investigation of non-interfacial residues led to the development of GADIS, which provides a method to adjust the residual structure of IDPs by shuffling the sequence order. Biophysical analysis of shuffled IDP sequences indicated that non-contacting residues can significantly alter the binding affinity. This may be an important consideration when designing peptide inhibitors.

9.6.3 Oligomerisation in the BCL-2 Family

Although they are not IDPs, the BCL-2-like proteins are important components of a signalling system that involves disordered proteins. These proteins are involved in apoptosis, with pro-apoptotic members oligomerising to form pores in the mitochondrial membrane. Biophysical characterisation of these apoptotic arbitrators indicated that both the pro-apoptotic and anti-apoptotic members could oligomerise, when the folded structure was destabilised with chemical denaturants. A spectrum of oligomerising ability was observed, which could provide clues to the functional evolution of these proteins.

9.6.4 Transition State Information is Encoded within the IDP

When stabilised by a partner macromolecule, IDPs can fold to a stable 3-dimensional structure. As the stability of this structure is dependent on the partner, it is possible that the partner macromolecule templates the folding reaction. In opposition to this hypothesis, comparative φ-value analyses revealed that the folding pathway for the IDP was found to be encoded within its sequence, rather than in the stabilising partner protein.

This produces an intriguing hypothesis relating to the functional advantage of disorder. If IDPs can encode both their transition states and their level of residual structure, it gives them the ability to evolve specific kinetic profiles. IDPs are typically involved in cell signalling, where responses to stimuli may have to occur quickly (e.g. activation of a cell surface receptor), or may need to be decisive and

relatively irreversible (e.g. stimulation of apoptosis). Changing the residual structure or encoded transition state provides an accessible method for evolution to tune the lifetimes of these complexes, which may be one explanation for the evolutionary conservation of disorder.

9.7 Future Directions

In terms of the work described in this thesis, the most promising piece of ongoing work is the oligomerisation of the BCL-2-like proteins. It would be fascinating to see whether the ancestor of the modern-day proteins was more, or less, prone to oligomerisation.

The comparative φ-value analyses provide another area of work that could be built on. Liza Dahal, a PhD student in Clarke Group, is currently investigating how φ-values for residues on the folded protein, MCL-1, compare when the IDP is changed. Secondly, it would be intriguing to study a IDR that can bind in different conformations (e.g. the C-terminal domain of p53 that was shown in Figure 1.8). How is the transition state encoded when the final structure is distinct? Comparative analysis provides a powerful framework to address these fundamental questions.

Bibliography

Ameisen, J.C. (2002). On the origin, evolution, and nature of programmed cell death: a timeline of four billion years. *Cell Death and Differentiation*, **9**, 367–393.

Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. *Science*, **181**, 223–230.

Anfinsen, C.B., Haber, E., M, S., White, F.H. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*, **47**, 1309–1314.

Avalos, J.L., Celic, I., Muhammad, S., Cosgrove, M.S., Boeke, J.D., Wolberger, C. (2002). Structure of a Sir2 enzyme bound to an acetylated p53 peptide. *Molecular Cell*, **10**, 523–535.

Aznauryan, M., Delgado, L., Soranno, A., Nettels, D., Huang, J., Labhardt, A.M., Grzesiek, S., Schuler, B. (2016). Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS. *Proceedings of the National Academy of Sciences*, **113**, E5389–E5398.

Babu, M.M., van der Lee, R., de Groot, N.S., Gsponer, J. (2011). Intrinsically disordered proteins: regulation and disease. *Current Opinion in Structural Biology*, **21**, 432–440.

Bah, A., Forman-Kay, J.D. (2016). Modulation of intrinsically disordered protein function by post-translational modifications. *Journal of Biological Chemistry*, **291**, 6696–6705.

Bah, A., Vernon, R.M., Siddiqui, Z., Krzeminski, M., Muhandiram, R., Zhao, C., Sonenberg, N., Kay, L.E., Forman-Kay, J.D. (2015). Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature*, **519**, 106–109.

Ball, P. (2011). More than a bystander. *Nature*, **478**, 467–468.

Barlow, D.J., Thornton, J.M. (1988). Helix geometry in proteins. *Journal of Molecular Biology*, **201**, 601–619.

Bartlett, G.J., Porter, C.T., Borkakoti, N., Thornton, J.M. (2002). Analysis of catalytic

residues in enzyme active sites. *Journal of Molecular Biology*, **324**, 105–121.

Berlow, R.B., Dyson, H.J., Wright, P.E. (2017). Hypersensitive termination of the hypoxic response by a disordered protein switch. *Nature*, **543**, 447–451.

Billings, K.S., Best, R.B., Rutherford, T.J., Clarke, J. (2008). Crosstalk between the Protein Surface and Hydrophobic Core in a Core-swapped Fibronectin Type III Domain. *Journal of Molecular Biology*, **375**, 560–571.

Binolfi, A., Theillet, F.-X., Selenko, P. (2012). Bacterial in-cell NMR of human α-synuclein: a disordered monomer by nature? *Biochemical Society Transactions*, **40**, 950–954.

Blommel, P.G., Fox, B.G. (2007). A combined approach to improving large-scale production of tobacco etch virus protease. *Protein Expression and Purification*, **55**, 53–68.

Bogner, C., Leber, B., Andrews, D.W. (2010). Apoptosis: Embedded in membranes. *Current Opinion in Cell Biology*, **22**, 845–851.

Borcherds, W., Theillet, F.-X., Katzer, A., Finzel, A., Mishall, K.M., Powell, A.T., Wu, H., Manieri, W., Dieterich, C., Selenko, P., Loewer, A., Daughdrill, G.W. (2014). Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. *Nature Chemical Biology*, **10**, 1000–1002.

Boyer, S., Biswas, D., Kumar Soshee, A., Scaramozzino, N., Nizak, C., Rivoire, O. (2016). Hierarchy and extremes in selections from pools of randomized proteins. *Proceedings of the National Academy of Sciences*, **113**, 3482–3487.

Brangwynne, C.P., Tompa, P., Pappu, R.V. (2015). Polymer physics of intracellular phase transitions. *Nature Physics*, **11**, 899–904.

Brosius, J., Erfle, M., Storella, J. (1985). Spacing of the -10 and -35 regions in the tac promoter. Effect on its in vivo activity. *Journal of Biological Chemistry*, **260**, 3539–3541.

Brown, C.J., Johnson, A.K., Dunker, A.K., Daughdrill, G.W. (2011). Evolution and disorder. *Current Opinion in Structural Biology*, **21**, 441–446.

Camilloni, C., Bonetti, D., Morrone, A., Giri, R., Dobson, C.M., Brunori, M., Gianni, S., Vendruscolo, M. (2016). Towards a structural biology of the hydrophobic effect in protein folding. *Scientific Reports*, **6**, 28285.

Camilloni, C., De Simone, A., Vranken, W.F., Vendruscolo, M. (2012). Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry*, **51**, 2224–2231.

Chen, J. (2012). Towards the physical basis of how intrinsic disorder mediates protein function. *Archives of Biochemistry and Biophysics*, **524**, 123–31.

Chen, J.W., Romero, P., Uversky, V.N., Dunker, A.K. (2006). Conservation of intrinsic disorder in protein domains and families: II. Functions of conserved disorder. *Journal of Proteome Research*, **5**, 888–898.

Chen, Y.-H., Yang, J.T. (1971). A new approach to the calculation of secondary structures of globular proteins by optical rotatory dispersion and circular dichroism. *Biochemical and Biophysical Research Communications*, **44**, 1285–1291.

Chen, Y.H., Yang, J.T., Chau, K.H. (1974). Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry*, **13**, 3350–3359.

Chi, X., Kale, J., Leber, B., Andrews, D.W. (2014). Regulating cell death at, on, and in membranes. *Biochimica et Biophysica Acta*, **1843**, 2100–2113.

Cho, J.-H., Raleigh, D.P. (2006). Denatured state effects and the origin of nonclassical phi values in protein folding. *Journal of the American Chemical Society*, **128**, 16492–16493.

Clapham, D.E. (2007). Calcium Signaling. *Cell*, **131**, 1047–1058.

Clarke, J., Hounslow, A.M., Bond, C.J., Fersht, A., Daggett, V. (2000). The effects of disulfide bonds on the denatured state of barnase. *Protein science : a publication of the Protein Society*, **9**, 2394–2404.

Crabtree, M.D., Borcherds, W., Poosapati, A., Shammas, S.L., Daughdrill, G.W., Clarke, J. (2017). Conserved Helix-Flanking Prolines Modulate Intrinsically Disordered Protein:Target Affinity by Altering the Lifetime of the Bound Complex. *Biochemistry*, **56**, 2379–2384.

Creamer, T.P., Rose, G.D. (1992). Side-chain entropy opposes a-helix formation but rationalizes experimentally determined helix-forming propensities (a-helix/protein folding/protein engineering). *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 5937–5941.

Czabotar, P.E., Lessene, G., Strasser, A., Adams, J.M. (2014). Control of apoptosis by

the BCL-2 protein family: implications for physiology and therapy. *Nature Reviews Molecular Cell Biology*, **15**, 49–63.

Dandliker, W.B., Hsu, M.L., Levin, J., Rao, B.R. (1981). Equilibrium and kinetic inhibition assays based upon fluorescence polarization. *Methods in Enzymology*, **74**, 3–28.

Das, R.K., Pappu, R. V (2013). Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences*, **110**, 13392–13397.

Day, C.L., Smits, C., Fan, F.C., Lee, E.F., Fairlie, W.D., Hinds, M.G. (2008). Structure of the BH3 domains from the p53-inducible BH3-only proteins Noxa and Puma in complex with Mcl-1. *Journal of Molecular Biology*, **380**, 958–971.

de los Rios, M. a, Muralidhara, B.K., Wildes, D., Sosnick, T.R., Marqusee, S., Wittung-Stafshede, P., Plaxco, K.W., Ruczinski, I. (2006). On the precision of experimentally determined protein folding rates and phi-values. *Protein science : a publication of the Protein Society*, **15**, 553–563.

Deber, M., Brandl, J., Hsu, L.C., Young, K., Deber, B. (1986). Amino Acid Composition of the Membrane and Aqueous of Integral Membrane Proteins. *Archives of Biochemistry and Biophysics*, **251**, 68–76.

Dempsey, G.T., Vaughan, J.C., Chen, K.H., Bates, M., Zhuang, X. (2011). Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nature Methods*, **8**, 1027–1036.

Dewson, G., Kratina, T., Czabotar, P., Day, C.L., Adams, J.M., Kluck, R.M. (2009). Bak Activation for Apoptosis Involves Oligomerization of Dimers via Their α6 Helices. *Molecular Cell*, **36**, 696–703.

Dewson, G., Kratina, T., Sim, H.W., Puthalakath, H., Adams, J.M., Colman, P.M., Kluck, R.M. (2008). To Trigger Apoptosis, Bak Exposes Its BH3 Domain and Homodimerizes via BH3:Groove Interactions. *Molecular Cell*, **30**, 369–380.

Dill, K.A. (1990). Dominant Forces in Protein Folding. *Biochemistry*, **39**, 7133–7155.

Dobson, C.M. (2003). Protein folding and misfolding. *Nature*, **426**, 884–890.

Dogan, J., Mu, X., Engström, Å., Jemth, P. (2013). The transition state structure for coupled binding and folding of disordered protein domains. *Scientific Reports*, **3**, 2076.

Donlan, R.M. (2001). Biofilm Formation: A Clinically Relevant Microbiological Process. *Clinical Infectious Diseases*, **33**, 1387–1392.

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C., Obradovic, Z. (2001). Intrinsically disordered protein. *Journal of Molecular Graphics & Modelling*, **19**, 26–59.

Dyson, H.J., Wright, P.E. (2002). Coupling of folding and binding for unstructured proteins. *Current opinion in structural biology*, **12**, 54–60.

Edwards, Y.J.K.J., Lobley, A.E., Pentony, M.M., Jones, D.T. (2009). Insights into the regulation of intrinsically disordered proteins in the human proteome by analyzing sequence and gene expression data. *Genome Biology*, **10**, R50.

Ellis, H., Horvitz, H. (1986). Genetic control of programmed cell death in the nematode Caenorhabditis elegans. *Cell*, **44**, 817–829.

Elmore, S. (2007). Apoptosis: A Review of Programmed Cell Death. *Toxicologic Pathology*, **35**, 495–516.

Fersht, A. (1999). Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding. W.H. Freeman.

Fersht, A.R., Matouschek, A., Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *Journal of Molecular Biology*, **224**, 771–782.

Fersht, A.R., Sato, S. (2004). Phi-value analysis and the nature of protein-folding transition states. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 7976–7981.

Fersht, A.R., Shi, J.-P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M.Y., Winter, G. (1985). Hydrogen bonding and biological specificity analyzed by protein engineering. *Nature*, **314**, 235–238.

Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft*, **27**, 2985–2993.

Fuxreiter, M., Simon, I., Friedrich, P., Tompa, P. (2004). Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *Journal of*

Molecular Biology, **338**, 1015–1026.

Fuxreiter, M., Tompa, P., Simon, I. (2007). Local structural disorder imparts plasticity on linear motifs. *Bioinformatics*, **23**, 950–956.

Germain, M., Duronio, V. (2007). The N terminus of the anti-apoptotic BCL-2 homologue MCL-1 regulates its localization and function. *Journal of Biological Chemistry*, **282**, 32233–32242.

Gianni, S., Dogan, J., Jemth, P. (2014). Distinguishing induced fit from conformational selection. *Biophysical Chemistry*, **189C**, 33–39.

Gianni, S., Jemth, P. (2014). Conserved nucleation sites reinforce the significance of Phi value analysis in protein-folding studies. *IUBMB Life*, **66**, 449–452.

Gill, S.C., von Hippel, P.H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry*, **182**, 319–326.

Giri, R., Morrone, A., Toto, A., Brunori, M., Gianni, S. (2013). Structure of the transition state for the binding of c-Myb and KIX highlights an unexpected order for a disordered system. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 14942–14947.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research*, **38**, W695–W699.

Greenfield, N., Fasman, G.D. (1969). Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry*, **8**, 4108–4116.

Gruszka, D.T., Mendonça, C.A.T.F., Paci, E., Whelan, F., Hawkhead, J., Potts, J.R., Clarke, J. (2016). Disorder drives cooperative folding in a multidomain protein. *Proceedings of the National Academy of Sciences*, **113**, 11841–11846.

Gruszka, D.T., Whelan, F., Farrance, O.E., Fung, H.K.H., Paci, E., Jeffries, C.M., Svergun, D.I., Baldock, C., Baumann, C.G., Brockwell, D.J., Potts, J.R., Clarke, J. (2015). Cooperative folding of intrinsically disordered domains drives assembly of a strong elongated protein. *Nature communications*, **6**, 7271.

Gruszka, D.T., Wojdyla, J.A., Bingham, R.J., Turkenburg, J.P., Manfield, I.W., Steward, A., Leech, A.P., Geoghegan, J.A., Foster, T.J., Clarke, J., Potts, J.R. (2012). Staphylococcal biofilm-forming protein has a contiguous rod-like structure.

Proceedings of the National Academy of Sciences, **109**, E1011–E1018.

Haq, S.R., Chi, C.N., Bach, A., Dogan, J., Engström, Å., Hultqvist, G., Karlsson, O.A., Lundström, P., Montemiglio, L.C., Strømgaard, K., Gianni, S., Jemth, P. (2012). Side-chain interactions form late and cooperatively in the binding reaction between disordered peptides and PDZ domains. *Journal of the American Chemical Society*, **134**, 599–605.

Harmon, T.S., Crabtree, M.D., Shammas, S.L., Posey, A.E., Clarke, J., Pappu, R. V (2016). GADIS: Algorithm for designing sequences to achieve target secondary structure profiles of intrinsically disordered proteins. *Protein Engineering, Design & Selection*, **29**, 339–46.

Hill, S.A., Kwa, L.G., Shammas, S.L., Lee, J.C., Clarke, J. (2014). Mechanism of assembly of the non-covalent spectrin tetramerization domain from intrinsically disordered partners. *Journal of Molecular Biology*, **426**, 21–35.

Hinds, M., Smits, C., Fredericks-Short, R., Risk, J., Bailey, M., Huang, D., Day, C. (2007). Bim, Bad and Bmf: intrinsically unstructured BH3-only proteins that undergo a localized conformational change upon binding to prosurvival Bcl-2 targets. *Cell Death and Differentiation*, **14**, 128–136.

Hofmeister, F. (1888). Zur Lehre von der Wirkung der Salze - Zweite Mittheilung **24**, 247–260.

Horovitz, A., Matthews, J.M., Fersht, A.R. (1992). Alpha-helix stability in proteins. II. Factors that influence stability at an internal position. *Journal of Molecular Biology*, **227**, 560–568.

Hunter, C.A. (2004). Quantifying intermolecular interactions: Guidelines for the molecular recognition toolbox. *Angewandte Chemie - International Edition*, **43**, 5310–5324.

Iešmantavičius, V., Dogan, J., Jemth, P., Teilum, K., Kjaergaard, M. (2014). Helical propensity in an intrinsically disordered protein accelerates ligand binding. *Angewandte Chemie - International Edition*, **53**, 1548–1551.

International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931–945.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.

Jenson, J.M., Ryan, J.A., Grant, R.A., Letai, A., Keating, A.E. (2017). Epistatic mutations in PUMA BH3 drive an alternate binding mode to potently and selectively inhibit anti-apoptotic Bfl-1. *eLife*, **6**, 1–23.

Johnson, B.A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods in molecular biology (Clifton, N.J.)*, **278**, 313–352.

Johnson, B.A., Blevins, R.A. (1994). NMR View: A computer program for the visualization and analysis of NMR data. *Journal of Biomolecular NMR*, **4**, 603–614.

Joseph, A.P., Srinivasan, N., De Brevern, A.G. (2012). Cis-trans peptide variations in structurally similar proteins. *Amino Acids*, **43**, 1369–1381.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., Phillips, D.C. (1958). A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature*, **181**, 662–666.

Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., G., H.R., Davies, D.R., Phillips, D.C., Shore, V.C. (1960). Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å. Resolution. *Nature*, **185**, 422–427.

Kerem, B., Rommens, J., Buchanan, J., Markiewicz, D., Cox, T., Chakravarti, A., Buchwald, M., Tsui, L. (1989). Identification of the cystic fibrosis gene: genetic analysis. *Science*, **245**, 1073–1080.

Keskin, O., Yuret, D., Gursoy, A., Turkay, M., Erman, B. (2004). Relationships between amino acid sequence and backbone torsion angle preferences. *Proteins: Structure, Function and Genetics*, **55**, 992–998.

Kostakioti, M., Hadjifrangiskou, M., Hultgren, S.J. (2013). Bacterial Biofilms: Development, Dispersal, and Therapeutic Strategies in the Dawn of the Postantibiotic Era. *Cold Spring Harbor Perspectives in Medicine*, a010306.

Krieger, J.M., Fusco, G., Lewitzky, M., Simister, P.C., Marchant, J., Camilloni, C., Feller, S.M., De Simone, A. (2014). Conformational recognition of an intrinsically disordered protein. *Biophysical journal*, **106**, 1771–1779.

Ku, B., Liang, C., Jung, J.U., Oh, B.-H. (2011). Evidence that inhibition of BAX activation by BCL-2 involves its tight and preferential interaction with the BH3 domain of BAX. *Cell Research*, **21**, 627–41.

Lau, Y.H., de Andrade, P., Wu, Y., Spring, D.R. (2015). Peptide stapling techniques

based on different macrocyclisation chemistries. *Chemical Society Reviews*, **44**, 91–102.

Lee, C., Kalmar, L., Xue, B., Tompa, P., Daughdrill, G.W., Uversky, V.N., Han, K.-H. (2014). Contribution of proline to the pre-structuring tendency of transient helical secondary structure elements in intrinsically disordered proteins. *Biochimica et Biophysica Acta*, **1840**, 993–1003.

Li, H., Zhu, H., Xu, C., Yuan, J. (1998). Cleavage of BID by Caspase 8 Mediates the Mitochondrial Damage in the Fas Pathway of Apoptosis. *Cell*, **94**, 491–501.

Li, J., Yuan, J. (2008). Caspases in apoptosis and beyond. *Oncogene*, **27**, 6194–6206.

Liu, Q., Moldoveanu, T., Sprules, T., Matta-Camacho, E., Mansur-Azzam, N., Gehring, K. (2010). Apoptotic regulation by MCL-1 through heterodimerization. *Journal of Biological Chemistry*, **285**, 19615–19624.

Lovell, S.C., Davis, I.W., Adrendall, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., Richardson, D.C. (2003). Structure validation by C alpha geometry: phi,psi and C beta deviation. *Proteins-Structure Function and Genetics*, **50**, 437–450.

Lowe, E.D., Tews, I., Cheng, K.Y., Brown, N.R., Gul, S., Noble, M.E.M., Gamblin, S.J., Johnson, L.N. (2002). Specificity determinants of recruitment peptides bound to phospho-CDK2/cyclin A. *Biochemistry*, **41**, 15625–15634.

Lum, K., Chandler, D., Weeks, J.D. (1999). Hydrophobicity at Small and Large length Scales. *Journal of Physical Chemistry B*, **103**, 4570–4577.

MacArthur, M.W., Thornton, J.M. (1991). Influence of proline residues on protein conformation. *Journal of Molecular Biology*, **218**, 397–412.

Markley, J.L., Bax, A., Arata, Y., Hilbers, C.W., Kaptein, R., Sykes, B.D., Wright, P.E., Wüthrich, K. (1998). Recommendations for the presentation of NMR structures of proteins and nucleic acids. *Journal of Biomolecular NMR*, **12**, 1–23.

Matouschek, A., Kellis, J.T., Serrano, L., Fersht, A.R. (1989). Mapping the transition state and pathway of protein folding by protein engineering. *Nature*, **340**, 122–126.

Maxwell, K.L., Wildes, D., Zarrine-Afsar, A., De Los Rios, M. a, Brown, A.G., Friel, C.T., Hedberg, L., Horng, J.-C., Bona, D., Miller, E.J., Vallée-Bélisle, A., Main, E.R.G.,

Bemporad, F., Qiu, L., Teilum, K., Vu, N.-D., Edwards, A.M., Ruczinski, I., Poulsen, F.M., Kragelund, B.B., Michnick, S.W., Chiti, F., Bai, Y., Hagen, S.J., Serrano, L., Oliveberg, M., Raleigh, D.P., Wittung-Stafshede, P., Radford, S.E., Jackson, S.E., Sosnick, T.R., Marqusee, S., Davidson, A.R., Plaxco, K.W. (2005). Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein science : a publication of the Protein Society*, **14**, 602–616.

McDonnell, J.M., Fushman, D., Milliman, C.L., Korsmeyer, S.J., Cowburn, D. (1999). Solution Structure of the Proapoptotic Molecule BID: A Structural Basis for Apoptotic Agonists and Antagonists. *Cell*, **96**, 625–634.

Milo, R. (2013). What is the total number of protein molecules per cell volume? A call to rethink some published values. *BioEssays*, **35**, 1050–1055.

Mirny, L. a, Shakhnovich, E.I. (1999). Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. *Journal of Molecular Biology*, **291**, 177–196.

Miroux, B., Walker, J.E. (1996). Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology*, **260**, 289–298.

Mitrea, D.M., Kriwacki, R.W. (2016). Phase separation in biology; functional organization of a higher order. *Cell Communication and Signaling*, **14**, 1.

Mittag, T., Forman-Kay, J.D. (2007). Atomic-level characterization of disordered protein ensembles. *Current Opinion in Structural Biology*, **17**, 3–14.

Morrisett, J.D., David, J.S., Pownall, H.J., Gotto, A.M. (1973). Interaction of an apolipoprotein (apoLP-alanine) with phosphatidylcholine. *Biochemistry*, **12**, 1290–1299.

Mujtaba, S., He, Y., Zeng, L., Yan, S., Plotnikova, O., Sachchidanand, Sanchez, R., Zeleznik-Le, N.J., Ronai, Z., Zhou, M.-M. (2004). Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. *Molecular Cell*, **13**, 251–63.

Muñoz, V., Serrano, L. (1995). Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. *Journal of Molecular Biology*, **245**, 297–308.

Myers, J.K., Pace, C.N., Scholtz, J.M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Science*, **4**, 2138–2148.

Naganathan, A.N., Muñoz, V. (2010). Insights into protein folding mechanisms from large scale analysis of mutational effects. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 8611–8616.

Nikolovska-Coleska, Z., Wang, R., Fang, X., Pan, H., Tomita, Y., Li, P., Roller, P.P., Krajewski, K., Saito, N.G., Stuckey, J.A., Wang, S. (2004). Development and optimization of a binding assay for the XIAP BIR3 domain using fluorescence polarization. *Analytical Biochemistry*, **332**, 261–273.

Nilsson, I., von Heijne, G. (1998). Breaking the camel's back: proline-induced turns in a model transmembrane helix. *Journal of Molecular Biology*, **284**, 1185–1189.

Nott, T.J., Craggs, T.D., Baldwin, A.J. (2016). Membraneless organelles can melt nucleic acid duplexes and act as biomolecular filters. *Nature Chemistry*, **8**, 569–575.

Nott, T.J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, D.P., Pawson, T., Forman-Kay, J.D., Baldwin, A.J. (2015). Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Molecular Cell*, **57**, 936–947.

O'Neil, K.T., DeGrado, W.F. (1990). A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science*, **250**, 646–51.

O'Sullivan, B.P., Freedman, S.D. (2009). Cystic fibrosis. *The Lancet*, **373**, 1891–1904.

Okur, H.I., Hladílková, J., Rembert, K.B., Cho, Y., Heyda, J., Dzubiella, J., Cremer, P.S., Jungwirth, P. (2017). Beyond the Hofmeister Series: Ion-Specific Effects on Proteins and Their Biological Functions. *Journal of Physical Chemistry B*, **121**, 1997–2014.

Onuchic, J.N., Wolynes, P.G. (2004). Theory of protein folding. *Current Opinion in Structural Biology*, **14**, 70–75.

Ou, L., Matthews, M., Pang, X., Zhou, H.X. (2017). The dock-and-coalesce mechanism for the association of a WASP disordered region with the Cdc42 GTPase. *FEBS Journal*, **284**, 3381–3391.

Ozenne, V., Schneider, R., Yao, M., Huang, J., Salmon, L., Zweckstetter, M., Jensen,

M.R., Blackledge, M. (2012). Mapping the Potential Energy Landscape of Intrinsically Disordered Proteins at Amino Acid Resolution. *Journal of the American Chemical Society*, **134**, 15138–15148.

Ozkan, S.B., Bahar, I., Dill, K. a (2001). Transition states and the meaning of Phi-values in protein folding kinetics. *Nature Structural Biology*, **8**, 765–769.

Pace, C.N., Scholtz, J.M., Grimsley, G.R. (2014). Forces stabilizing proteins. *FEBS Letters*, **588**, 2177–2184.

Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Science*, **4**, 2411–2423.

Papadakos, G., Sharma, A., Lancaster, L.E., Bowen, R., Kaminska, R., Leech, A.P., Walker, D., Redfield, C., Kleanthous, C. (2015). Consequences of inducing intrinsic disorder in a high-affinity protein-protein interaction. *Journal of the American Chemical Society*, **137**, 5252–5255.

Pappu, R. V, Srinivasan, R., Rose, G.D. (2000). The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 12565–12570.

Peng, Z., Mizianty, M.J., Kurgan, L. (2013). Genome-scale prediction of proteins with long intrinsically disordered regions. *Proteins*, 1–14.

Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, **7**, 95–99.

Richardson, J.S., Richardson, D.C. (1988). Amino acid preferences for specific locations at the ends of alpha helices. *Science (New York, N.Y.)*, **240**, 1648–52.

Riordan, J., Rommens, J., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J., Al., E. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*, **245**, 1066–1073.

Rogers, J.M., Oleinikovas, V., Shammas, S.L., Wong, C.T., De Sancho, D., Baker, C.M., Clarke, J. (2014a). Interplay between partner and ligand facilitates the folding and binding of an intrinsically disordered protein. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 15420–15425.

Rogers, J.M., Wong, C.T., Clarke, J. (2014b). Coupled folding and binding of the disordered protein PUMA does not require particular residual structure. *Journal of the American Chemical Society*, **136**, 5197–5200.

Rogers, J.M.J., Steward, A., Clarke, J. (2013). Folding and binding of an intrinsically disordered protein: fast, but not "diffusion-limited". *Journal of the American Chemical Society*, **135**, 1415–1422.

Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., Dunker, A.K. (2001). Sequence complexity of disordered protein. *Proteins*, **42**, 38–48.

Rommens, J., Iannuzzi, M., Kerem, B., Drumm, M., Melmer, G., Dean, M., Rozmahel, R., Cole, J., Kennedy, D., Hidaka, N., Al., E. (1989). Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science*, **245**, 1059–1065.

Rosano, G.L., Ceccarelli, E. a. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology*, **5**, 1–17.

Roth, C.M., Neal, B.L., Lenhoff, A.M. (1996). Van der Waals Interactions Involving Proteins. *Biophysical Journal*, **70**, 977–987.

Rustandi, R.R., Baldisseri, D.M., Weber, D.J. (2000). Structure of the negative regulatory domain of p53 bound to S100B(ββ). *Nature structural biology*, **7**, 570–574.

Sánchez, I.E., Kiefhaber, T. (2003). Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *Journal of Molecular Biology*, **325**, 367–376.

Schlessinger, A., Schaefer, C., Vicedo, E., Schmidberger, M., Punta, M., Rost, B. (2011). Protein disorder — a breakthrough invention of evolution? *Current opinion in Structural Biology*, **21**, 412–8.

Schreiber, G., Fersht, A.R. (1996). Rapid, electrostatically assisted association of proteins. *Nature Structural Biology*, **3**, 427–31.

Serrano, L. (1995). Comparison between the φ-distribution of the amino acids in the Protein Data Base and NMR data indicates that amino acids have various φ propensities in the random coil conformation. *Journal of Molecular Biology*, **254**, 322–333.

Shamas-din, A., Kale, J., Leber, B., Andrews, D.W. (2013). Mechanisms of Action of

Bcl-2 Family Proteins. *Cold Spring Harbor Perspectives in Biology*, **5**, 1–22.

Shammas, S.L. (2017). Mechanistic roles of protein disorder within transcription. *Current Opinion in Structural Biology*, **42**, 155–161.

Shammas, S.L., Crabtree, M.D., Dahal, L., Wicky, B.I.M., Clarke, J. (2016). Insights into coupled folding and binding mechanisms from kinetic studies. *Journal of Biological Chemistry*, **291**, 6689–6695.

Shammas, S.L., Rogers, J.M., Hill, S.A., Clarke, J. (2012). Slow, Reversible, Coupled Folding and Binding of the Spectrin Tetramerization Domain. *Biophysical Journal*, **103**, 2203–2214.

Shammas, S.L., Travis, A.J., Clarke, J. (2014). Allostery within a transcription coactivator is predominantly mediated through dissociation rate constants. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 12055–12060.

Shammas, S.L.S.S.L., Travis, A.A.J.A.A.J., Clarke, J. (2013). Remarkably Fast Coupled Folding and Binding of the Intrinsically Disordered Transactivation Domain of cMyb to CBP KIX. *Journal of Physical Chemistry B*, **117**, 13346–13356.

Shih, O., Yeh, Y.-Q., Liao, K.-F., Sung, T.-C., Chiang, Y.-W., Jeng, U.-S. (2017). Oligomerization process of Bcl-2 associated X protein revealed from intermediate structures in solution. *Physical Chemistry Chemical Physics*, **19**, 7947–7954.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., Higgins, D.G., Aniba, M., Poch, O., Thompson, J., Blackshields, G., Sievers, F., Shi, W., Wilm, A., Higgins, D., Bradley, R., Roberts, A., Smoot, M., Juvekar, S., Do, J., Dewey, C., Holmes, I., Pachter, L., Clamp, M., Cuff, J., Searle, S., Barton, G., Do, C., Mahabhashyam, M., Brudno, M., Batzoglou, S., Eddy, S., Edgar, R., Edgar, R., Finn, R., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J., Gavin, O., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E., Eddy, S., Bateman, A., Gouy, M., Guindon, S., Gascuel, O., Hogeweg, P., Hesper, B., Katoh, K., Misawa, K., Kuma, K., Miyata, T., Katoh, K., Toh, H., Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., Lopez, R., Thompson, J., Gibson, T., Higgins, D., Lassmann, T., Sonnhammer, E., Liu, Y., Schmidt, B., Maskell, D., Löytynoja, A., Goldman, N., Mizuguchi, K., Deane, C., Blundell, T., Overington, J., Morgenstern, B., Frech, K., Dress, A., Werner, T., Notredame, C., Higgins, D., Heringa, J., Pirovano, W., Feenstra, K., Heringa, J., Söding, J., Thompson, J.,

Koehl, P., Ripp, R., Poch, O., Wilm, A., Higgins, D., Notredame, C. (2014). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, **7**, 539–539.

Sjöback, R., Nygren, J., Kubista, M. (1995). Absorption and Fluorescence Properties of Fluorescein. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **51**, L7–L21.

Smits, C., Czabotar, P.E., Hinds, M.G., Day, C.L. (2008). Structural plasticity underpins promiscuous binding of the prosurvival protein A1. *Structure*, **16**, 818–829.

Sobolev, V., Eyal, E., Gerzon, S., Potapov, V., Babor, M., Prilusky, J., Edelman, M. (2005). SPACE: A suite of tools for protein structure prediction and analysis based on complementarity and environment. *Nucleic Acids Research*, **33**, W39–W43.

Srinivasan, R., Rose, G.D. (1999). A physical basis for protein secondary structure. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 14258–14263.

Stewart, D.E., Sarkar, A., Wampler, J.E. (1990). Occurrence and role of cis peptide bonds in protein structures. *Journal of molecular biology*, **214**, 253–60.

Suchyna, T.M., Xu, L.X., Gao, F., Fourtner, C.R., Nicholson, B.J. (1993). Identification of a proline residue as a transduction element involved in voltage gating of gap junctions. *Nature*, **365**, 847–849.

Tamiola, K., Acar, B., Mulder, F.A.A. (2010). Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *Journal of the American Chemical Society*, **132**, 18000–18003.

Theillet, F.-X., Binolfi, A., Bekei, B., Martorana, A., Rose, H.M., Stuiver, M., Verzini, S., Lorenz, D., van Rossum, M., Goldfarb, D., Selenko, P. (2016). Structural disorder of monomeric α-synuclein persists in mammalian cells. *Nature*, **530**, 45–50.

Theillet, F.-X., Kalmar, L., Tompa, P., Han, K.-H., Selenko, P., Dunker, A.K., Daughdrill, G.W., Uversky, V.N. (2014). The alphabet of intrinsic disorder. *Intrinsically Disordered Proteins*, **1**, e24360–e24373.

Thomas, L.W., Lam, C., Edwards, S.W. (2010). Mcl-1; the molecular regulation of protein function. *FEBS Letters*, **584**, 2981–2989.

Thornton, J.W. (2004). Resurrecting ancient genes: experimental analysis of extinct

molecules. *Nature Reviews Genetics*, **5**, 366–375.

Tissot, A.C., Vuilleumier, S., Fersht, A.R. (1996). Importance of two buried salt bridges in the stability and folding pathway of barnase. *Biochemistry*, **35**, 6786–6794.

Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L., Tawfik, D.S. (2007). The Stability Effects of Protein Mutations Appear to be Universally Distributed. *Journal of Molecular Biology*, **369**, 1318–1332.

Tompa, P., Prilusky, J., Silman, I., Sussman, J.L. (2008). Structural disorder serves as a weak signal for intracellular protein degradation. *Proteins: Structure, Function and Genetics*, **71**, 903–909.

Toto, A., Camilloni, C., Giri, R., Brunori, M., Vendruscolo, M., Gianni, S. (2016). Molecular recognition by templated folding of an intrinsically disordered protein. *Scientific Reports*, **6**, 21994.

Toto, A., Gianni, S. (2016). Mutational analysis of the binding-induced folding reaction of the mixed-lineage leukemia protein to the KIX domain. *Biochemistry*, **55**, 3957–3962.

Toto, A., Giri, R., Brunori, M., Gianni, S. (2014). The mechanism of binding of the KIX domain to the mixed lineage leukemia protein and its allosteric role in the recognition of c-Myb. *Protein Science*, **23**, 962–969.

Towse, C.L., Vymetal, J., Vondrasek, J., Daggett, V. (2016). Insights into Unfolded Proteins from the Intrinsic φ/ψ Propensities of the AAXAA Host-Guest Series. *Biophysical Journal*, **110**, 348–361.

Uhlen, M., Fagerberg, L., Hallstrom, B., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C., Odeberg, J., Djureinovic, D., Takanen, J., Hober, S., Alm, T., Edqvist, P.-H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J., Ponten, F. (2015). Tissue-based map of the human proteome. *Science*, **347**, 1260419–1260419.

Uversky, V.N., Gillespie, J.R., Fink, A.L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*, **41**, 415–27.

Uversky, V.N., Oldfield, C.J., Dunker, a K. (2008). Intrinsically Disordered Proteins in Human Diseases: Introducing the D 2 Concept. *Annual Review of Biophysics*, **37**,

215–246.

Varadi, M., Kosol, S., Lebrun, P., Valentini, E., Blackledge, M., Dunker, A.K., Felli, I.C., Forman-Kay, J.D., Kriwacki, R.W., Pierattelli, R., Sussman, J., Svergun, D.I., Uversky, V.N., Vendruscolo, M., Wishart, D., Wright, P.E., Tompa, P. (2014). PE-DB: A database of structural ensembles of intrinsically disordered and of unfolded proteins. *Nucleic Acids Research*, **42**, 326–335.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V, Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigó, R., Campbell, M.J., Sjolander, K. V, Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A.,

Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X. (2001). The Sequence of the Human Genome. *Science*, **291**, 1304–1351.

Vitalis, A., Pappu, R. V. (2010). ABSINTH: A New Continuum Solvation Model for Simulations of Polypeptides in Aqueous Solutions. *Journal of Computational Chemistry*, **30**, 673–699.

Walker, D., Moore, G.R., James, R., Kleanthous, C. (2003). Thermodynamic consequences of bipartite immunity protein binding to the ribosomal ribonuclease colicin E3. *Biochemistry*, **42**, 4161–4171.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of Molecular Biology*, **337**, 635–645.

Wei, M.C., Lindsten, T., Mootha, V.K., Weiler, S., Gross, A., Ashiya, M., Thompson, C.B., Korsmeyer, S.J. (2000). tBID, a membrane-targeted death ligand, oligomerizes BAK to release cytochrome c. *Genes and Development*, **14**, 2060–2071.

Wensley, B.G., Kwa, L.G., Shammas, S.L., Rogers, J.M., Clarke, J. (2012). Protein folding: Adding a nucleus to guide helix docking reduces landscape roughness. *Journal of Molecular Biology*, **423**, 273–283.

Wicky, B.I.M., Shammas, S.L., Clarke, J. (2017). Affinity of IDPs to their targets is modulated by ion-specific changes in kinetics and residual structure. *Proceedings of the National Academy of Sciences*, **114**, 9882–9887.

Williams, K.A., Deber, C.M. (1991). Proline Residues in Transmembrane Helices : Structural or Dynamic Role ? *Biochemistry*, **30**, 8919–8923.

Wishart, D.S., Sykes, B.D. (1994). The 13C Chemical-Shift Index: A simple method for the identification of protein secondary structure using 13C chemical-shift data. *Journal of Biomolecular NMR*, **4**, 171–180.

Wright, P.E., Dyson, H.J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*, **293**, 321–331.

Yue, P., Li, Z., Moult, J. (2005). Loss of protein structure stability as a major causative factor in monogenic disease. *Journal of Molecular Biology*, **353**, 459–473.

Zaman, M.H., Shen, M.Y., Berry, R.S., Freed, K.F., Sosnick, T.R. (2003). Investigations into sequence and conformational dependence of backbone entropy, inter-basin dynamics and the flory isolated-pair hypothesis for peptides. *Journal of Molecular Biology*, **331**, 693–711.

Zhou, H.-X. (2012). Intrinsic disorder: signaling via highly specific but short-lived association. *Trends in Biochemical Sciences*, **37**, 43–48.

Zhou, H.-X., Pang, X., Lu, C. (2012). Rate constants and mechanisms of intrinsically disordered proteins binding to structured targets. *Physical chemistry chemical physics : PCCP*, **14**, 10466–10476.

Zwanzig, R., Szabo, A., Bagchi, B. (1992). Levinthal' s paradox. *Proceedings of the National Academy of Sciences*, **89**, 20–22.